MULTI-MODALITY PLASMA-BASED DETECTION OF MINIMAL RESIDUAL DISEASE IN TRIPLE-NEGATIVE BREAST CANCER

Yu-Hsiang Chen

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Medical and Molecular Genetics,
Indiana University

July 2019

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____
Milan Radovich, PhD, Chair

_____
Mircea Ivan, MD, PhD

April 30, 2019

_____
Lang Li, PhD

_____
Yunlong Liu, PhD

_____
Bryan P. Schneider, MD

_____
Todd C. Skaar, PhD

ACKNOWLEDGEMENT

It's been a long, but fantastic journey to be able to accomplish my PhD work in such a wonderful lab. I couldn't have done this without the help and support from so many of you.

First, I would like to thank my mentor Dr. Milan Radovich. I was so fortunate to be able to join your lab in 2014. You have always inspired me, not only on the scientific projects, but also in my career development. You have also provided many opportunities for me to grow, such as attending conferences and workshops that helped me broaden my horizons. Because of you, I had my first first-author publication, won the Scholar-in-Training award at AACR, and obtained a fantastic position in industry before graduation. I can't thank you enough for all you have provided me with during these 6 years.

To my lab mates, Jeffrey Solzak, Brad Hancock, Robin Paul, and Susanna Scott. I couldn't have done this without you guys. I was very privileged to be surrounded by a group of scientists with great rapport and also extremely helpful in many aspects. A special thanks to Jeffrey Solzak and Brad Hancock. We spent the most of the time of my PhD study together. We talked about science and life and you guys always put things in perspective. You two are more than colleagues, you are my brothers. Please remember you have a Taiwanese brother.

To my committee, Drs. Bryan Schneider, Lang Li, Yunlong Liu, Mircea Ivan, Todd Skaar, thank you for your time and suggestions on my projects. You have helped me to become a better scientist.

To our 2013 IBMG cohort, especially Xiaoting Wang, Lakshmi Prabhu, and Emma Longworth-Mills, thank you for being around whenever I needed help or someone to talk to regarding graduate work.

To my family back in Taiwan, you are my biggest moral support and you have always believed in me. I would also like to thank my grandma. When I was struggling, I always thought of you. I wish you could have seen what I have accomplished.

Thank you to every one of you that I may have forgotten who has been supportive during the past 6 years. I couldn't have done this work without any of your help.

Thank you!

Yu-Hsiang Chen

MULTI-MODALITY PLASMA-BASED DETECTION OF MINIMAL RESIDUAL DISEASE IN TRIPLE-NEGATIVE BREAST CANCER

Triple-negative breast cancers (TNBCs) are pathologically defined by the absence of estrogen, progesterone, and HER2 receptors. Compared to other breast cancers, TNBC has a relatively high mortality. In addition, TNBC patients are more likely to relapse in the first few years after treatment, and experiencing a shorter median time from recurrence to death. Detecting the presence of tumor in patients who are technically "disease-free" after neoadjuvant chemotherapy and surgery as early as possible might be able to predict recurrence of patients, and then provide timely intervention for additional therapy. To this end, I applied the analysis of "liquid biopsies" for early detection of minimal residual disease (MRD) on early-stage TNBC patients using next-generation sequencing. For the first part of this study, I focused on detecting circulating tumor DNA (ctDNA) from TNBC patients after neoadjuvant chemotherapy and surgery. First, patient-specific somatic mutations were identified by sequencing primary tumors. From these data, 82% of the patients had at least one TP53 mutation, followed by 16% of the patients having at least one PIK3CA mutation. Next, I sequenced matched plasma samples collected after surgery to identify ctDNA with the same mutations. I observed that by detecting corresponding ctDNA I was able to predict rapid recurrence, but not distant recurrence. To increase the sensitivity of MRD detection, in the second part I developed a strategy to co-detect ctDNA along with circulating tumor RNA (ctRNA). An advantage of ctRNA is its active release into the circulation from living cancer cells. Preliminary data showed that more mutant molecules were identified after incorporating ctRNA with

ctDNA detection in a metastatic breast cancer setting. A validation study in early-stage TNBC is in progress. In summary, my study suggests that co-detection of ctDNA and ctRNA could be a potential solution for the early detection of disease recurrence.

Milan Radovich, PhD, Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AF	Allele frequency
BAM	Binary alignment/mapping
BRCA	BReastCAncer gene
CA	Cancer antigen
CAPP-seq	CAncer Personalized Profiling by deep Sequencing
CAT	Computerized Axial Tomography.
cDNA	Complementary deoxyribonucleic acid
CEA	Carcinoembryonic antigen
cfDNA	Cell-free deoxyribonucleic acid
cfRNA	Cell-free ribonucleic acid
CI	Confidence interval
CONSORT	Consolidated Standards for Reporting Trials
CRISPR	Clustered regularly interspaced short palindromic repeats
CSF	Cerebrospinal fluid
CTC	Circulating tumor cell
ctDNA	Circulating tumor deoxyribonucleic acid
ctRNA	Circulating tumor ribonucleic acid
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddPCR	Droplet digital polymerase chain reaction
DFS	Disease-free survival
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dTTP	Deoxythymidine triphosphate
dPCR	Digital polymerase chain reaction
ECOG	Eastern Cooperative Oncology Group
EpCAM	Epithelial cell adhesion molecule
ER	Estrogen receptor
exoRNA	Exosomal ribonucleic acid

FDA	Food and Drug Administration
FFPE	Formalin-fixed, paraffin-embedded
FISH	Fluorescence in situ hybridization
HER2	Human epidermal growth factor receptor 2
H&E staining	Haemotoxylin and Eosin staining
HR	Hazard ratio
IGV	Integrative genomics viewer
IHC	Immunohistochemistry
IRB	Institutional Review Board
lncRNA	Long non-coding ribonucleic acid
microRNA	micro ribonucleic acid
MRD	Minimal residual disease
MRI	Magnetic resonance imaging
mRNA	messenger ribonucleic acid
NIPT	Noninvasive prenatal test
NGS	Next-generation sequencing
OS	Overall survival
Pap smear	Papanicolaou smear
PARP	Poly-ADP(ribose) polymerase
pCR	Pathological complete response
PCR	Polymerase chain reaction
PD-L1	Programmed death-ligand 1
PR	Progesterone receptor
PSA	Prostate-specific antigen
qPCR	Quantitative polymerase chain reaction
RCB	Residual cancer burden
RD	Residual disease
RNA	Ribonucleic acid
RT	Reverse transcriptase
TCGA	The Cancer Genome Atlas
TNBC	Triple-negative breast cancer

WES	Whole exome sequencing
WGS	Whole genome sequencing
WTS	Whole transcriptome sequencing

Chapter 1: Introduction

1.1 Triple-negative breast cancer

1.1.1 Clinical features and current standard of care of triple-negative breast cancer

Breast cancer is the most prevalent cancer type among women in the world and in the United States. In 2018, there were more than 260,000 estimated new cases of breast cancer and more than 40,000 are estimated to die from the disease [1]. Among those cases, about 15% of cases are categorized as triple-negative breast cancer (TNBC). TNBC is defined by the lack of estrogen (ER), progesterone (PR), and human epidermal growth factor receptor 2 (HER2) over-expression (Figure 1) [2-6]. While TNBC comprises a minority of breast cancer cases, it results in a disproportionally higher rate of mortality (Figure 2) [7]. Compared to ER- and HER2-positive disease, TNBCs have a higher incidence of visceral metastasis, a higher chance of recurrence within the first 3 years after chemotherapy and surgery, and a shorter overall survival (OS) after the onset of metastatic disease (Figure 3) [8, 9]. Compared to other subtypes of breast cancer, TNBCs tends to occur more often in women who are younger, of African-American/African descent, or Hispanic ethnicity. In addition, women who are carriers of pathologic germline mutations in BRCA1 have a higher risk of developing TNBC [10].

Histologically, TNBCs can be categorized into several subtypes. Around 95% of TNBCs are classified as invasive ductal carcinoma, followed by 1-2% as invasive lobular carcinoma, less than 1% as spindle-cell metaplastic carcinoma, adenoid cystic carcinoma, and other rare histologies. Among those, adenoid cystic carcinoma, adenosquamous

carcinoma and fibromatosis-like spindle-cell metaplastic carcinomas are less aggressive and usually relapse locally [11].

Unlike ER/PR positive or HER-2 positive breast cancers, who have endocrine therapy or HER-2 targeted therapy, patients with TNBC typically receive chemotherapy alone. The current standard of care for TNBC patients incorporates the use of various chemotherapy including anthracyclines (e.g. doxorubicin), taxanes (e.g. paclitaxel and docetaxel), and cyclophosphamide in the adjuvant and neoadjuvant settings. More recently, PARP inhibition has been FDA approved for HER2-negative breast cancers in patients that carry a deleterious BRCA1/2 germline mutation [12, 13]. And in March 2019, the FDA approved the first immunotherapy for metastatic TNBC, atezolizumab, for patients whose tumors have a PD-L1 IHC staining of 1% or greater based on results from the IMpassion130 trial [14].

A significant proportion of patients with early-stage TNBC are treated with neoadjuvant chemotherapy. Based on response of patients to neoadjuvant therapy, a stark dichotomy exists in outcome. Approximately a third of patients will achieve a pathologic complete response (pCR) and go on to have a favorable overall survival (OS) outcome (94% at 3 years). On the contrast, two-thirds of patients have residual disease after neoadjuvant chemotherapy and are at a high risk of relapse leading to an inferior OS (68% at 3 years) [15].

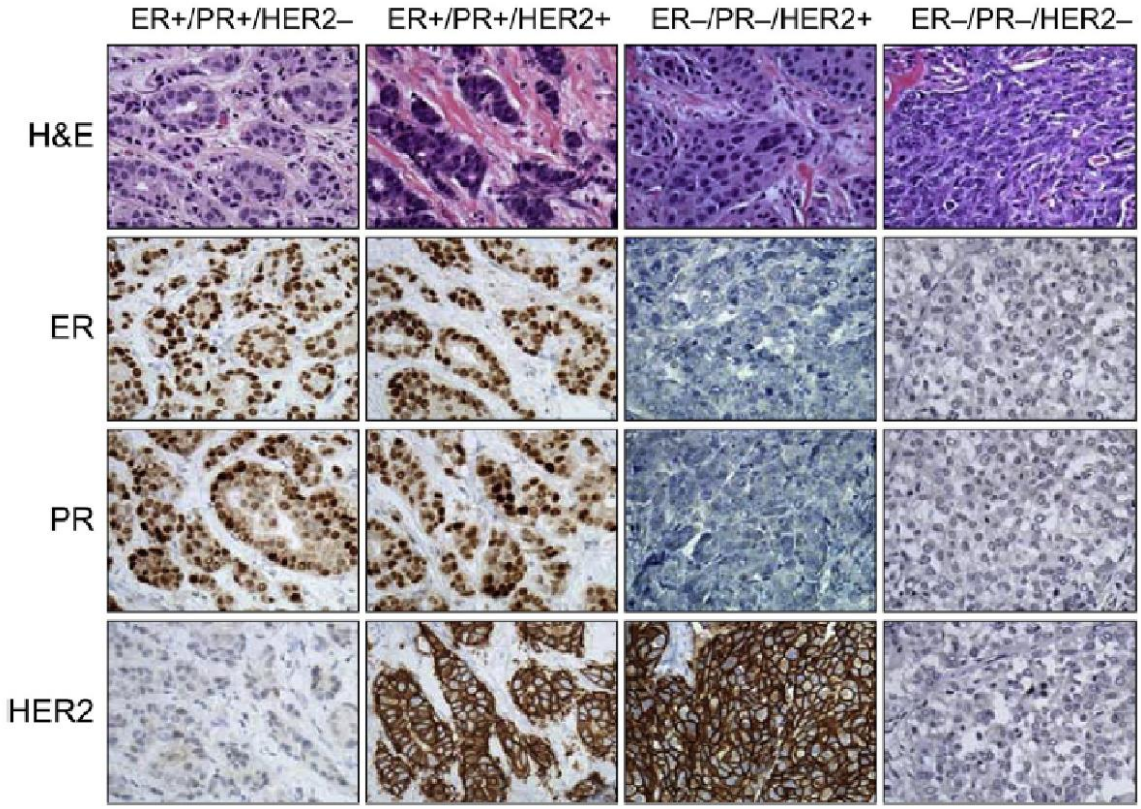


Figure 1: The pathological patterns of different subtypes of breast cancer. H&E staining shows cancer histology. The positive or negative status for hormone receptors ER, PR and HER2 categorizes breast cancers into four major clinical subtypes. Adapted from Rivenbark et al. [5].

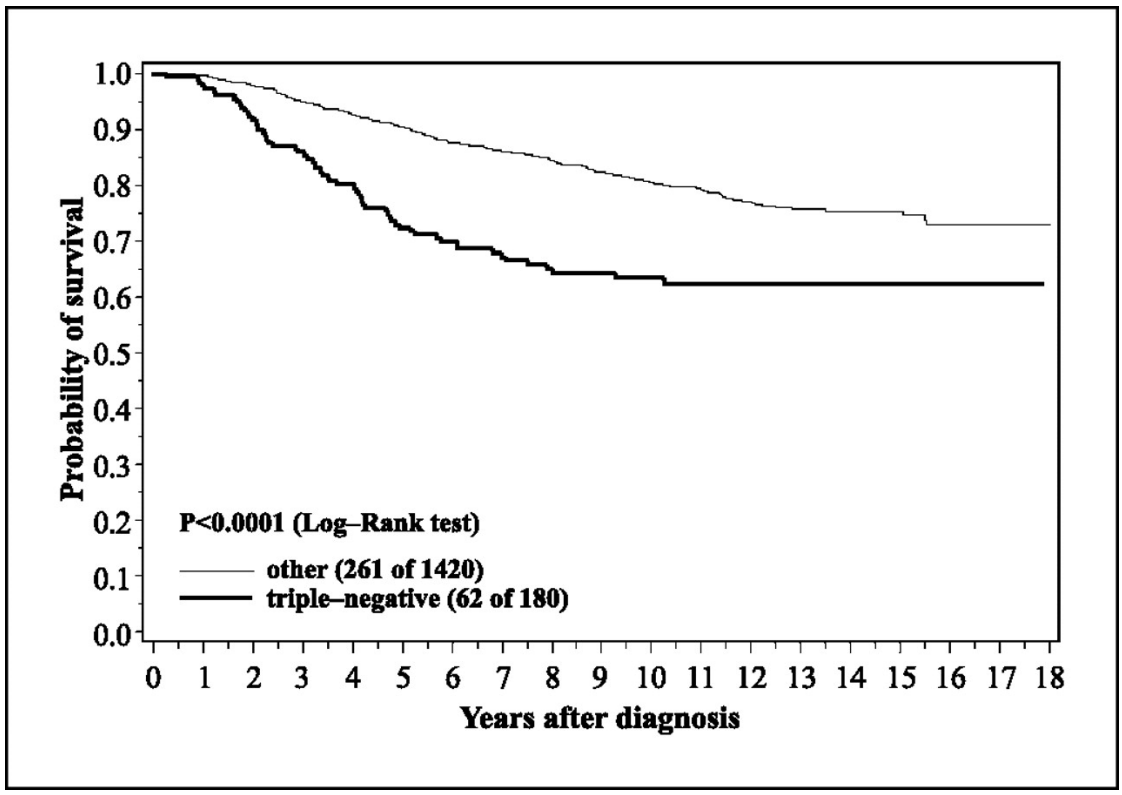


Figure 2: Rates of breast-specific survival in triple-negative and other breast cancers.

Adpated from Dent et al. [7].

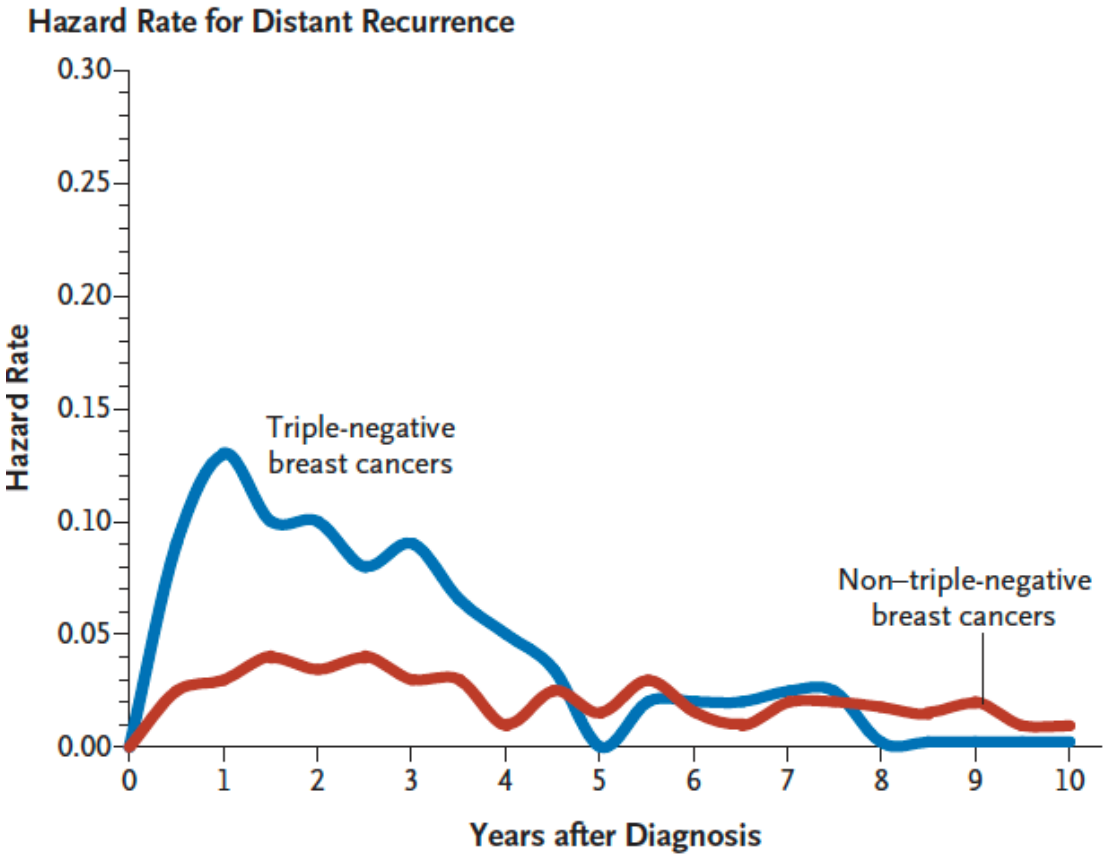


Figure 3: The hazard rates of distant recurrence of triple-negative breast cancer and non-triple-negative breast cancer. Triple-negative breast cancer is more likely to relapse in the first three years after treatment. Adapted from Foulkes et al. [10].

1.1.2 Molecular characteristics of TNBC

TNBCs are a collection of breast tumors that are negative for ER, PR and HER-2 by immunohistochemistry (IHC) and/or fluorescence in situ hybridization (FISH). Although clinical guidance for breast cancer patients still depend mainly on those three markers, advancements in 'omics' technologies over the past two decades have allowed us to have a better understanding of this disease on the molecular level, and perhaps to discover molecular patterns that would have clinical benefits for subgroups of patients. In 2000, Perou et al. had utilized cDNA microarrays representing 8102 human genes to characterize gene expression patterns of human breast tumors [16]. In this study, breast cancers were classified as HER-2, Luminal A and B, normal-like, and basal-like (Figure 4) [17, 18]. Further studies showed an association between basal-like breast cancer and germline BRCA1 mutations [19-21].

Basal-like breast cancer is often referred to as TNBC since the majority of the basal-like tumors are usually negative for ER, PR and HER-2. However, not all TNBC tumors are basal-like breast cancer: around 50% to 75% of basal-like breast cancers are classified as TNBC [22]. Conversely, basal-like breast cancer represents the most prevalent subtype of TNBC (55% to 81%) (Figure 5) [23, 24].

BRCA1 mutations are found in less than 5% in sporadic tumors [25]. However, women with the presence of germline mutation in BRCA1 will have higher risk to develop breast cancer lifetime [26]. There is a strong association existing between BRCA1 mutations, TNBC and basal-like phenotype. More than 75% of breast cancer tumors carrying BRCA1 mutations have a basal-like phenotype, TNBC phenotype, or both [27, 28].

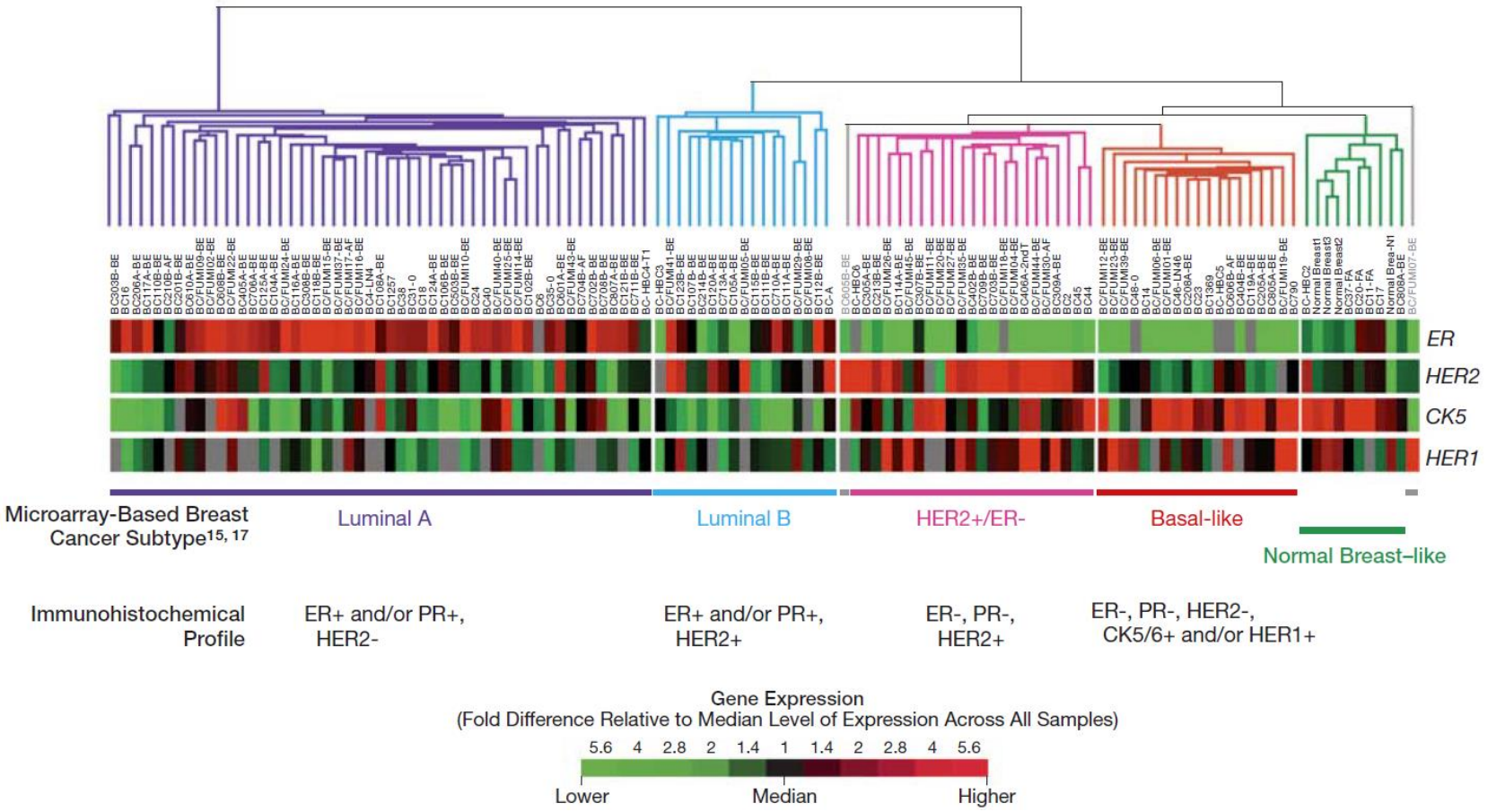


Figure 4: Example of stratification of breast cancer using gene expression patterns for hierarchical cluster analysis. Breast tumors are clustered into five intrinsic subtypes: Luminal A and B, HER2, Basal-like, and Normal Breast-like. Adapted from Carey et al. [18].

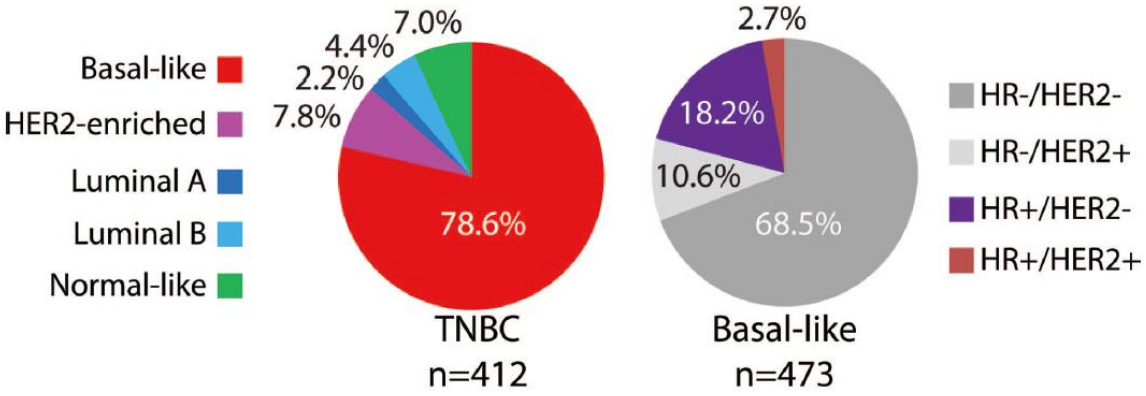


Figure 5: Comparison of intrinsic molecular subtypes and pathological subtypes among TNBC tumors and basal-like breast cancer tumor. The most frequent intrinsic subtype of TNBC is basal-like; the most frequent pathological subtype of basal-like breast cancer is TNBC. HR: hormone receptor. Adapted from Prat et al. [24].

1.2 Liquid biopsy on cancer detection and therapeutic monitoring

1.2.1 Sources of liquid biopsy for cancer detection

Tissue biopsy samples are commonly used in the clinic to diagnose, histologically characterize, and determine the hormonal status of breast tumors. Further, the tumor-derived material from biopsy can potentially provide pathological and molecular information to guide the selection of therapy. However, there are still limitations of utilizing tissue biopsy. Some lesions are not safe or feasible to access. In addition, it is also difficult to perform serial sampling on tissue samples. Moreover, tissue biopsy samples cannot completely represent the whole tumor bulk and may not capture the heterogeneity for a given patient's disease.

An emerging method for non-invasive cancer detection is the analysis of tumor-derived material from body fluids, such as blood, also known as "liquid biopsies". Compared to tissue biopsy, liquid biopsy is less invasive and more feasible for patients. Liquid biopsy can capture the mutational heterogeneity of a patient's disease, in particular for those with multi-focal metastatic disease, where cell-free DNA from the various metastases is shed into the circulation and analyzed. Furthermore, by serial sampling, liquid biopsy can be utilized to monitor treatment response, to potentially discover mutations which will lead to therapeutic resistance [29] (Figure 6), and also to detect minimal residual disease (MRD) prior to clinical recurrence.

Although the majority of the studies on liquid biopsy have focused on blood or plasma, other bodily fluids have been selected according to tumor types, such as urine for bladder cancer and prostate cancer, saliva for head and neck cancer, and cerebrospinal

fluid (CSF) for brain cancer and neurodegenerative disease [30] (Figure 7). Other sources of non-invasive biopsy have also been used to detect the presence of tumor, including: Pap smear, pancreatic juice, stool, pleural effusions, and ascites etc [29, 30].

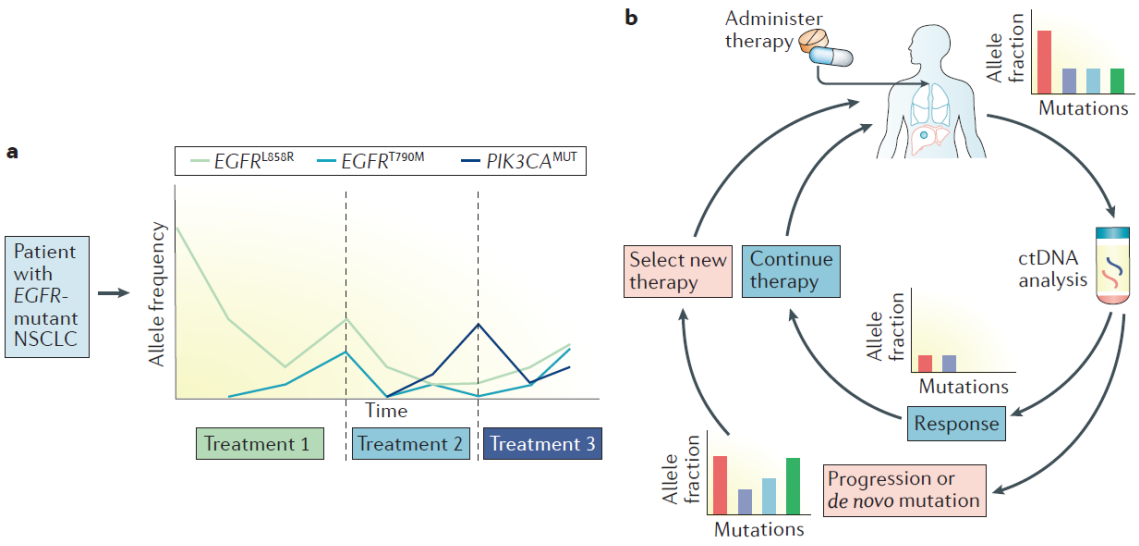


Figure 6: Liquid biopsy can be used to monitor the response of treatment. Serial sampling of plasma to detect different mutations over times and help guide the selection of treatment: continue the same therapy or select a new therapy according to the mutations identified. Adapted from Wan et al. [29].

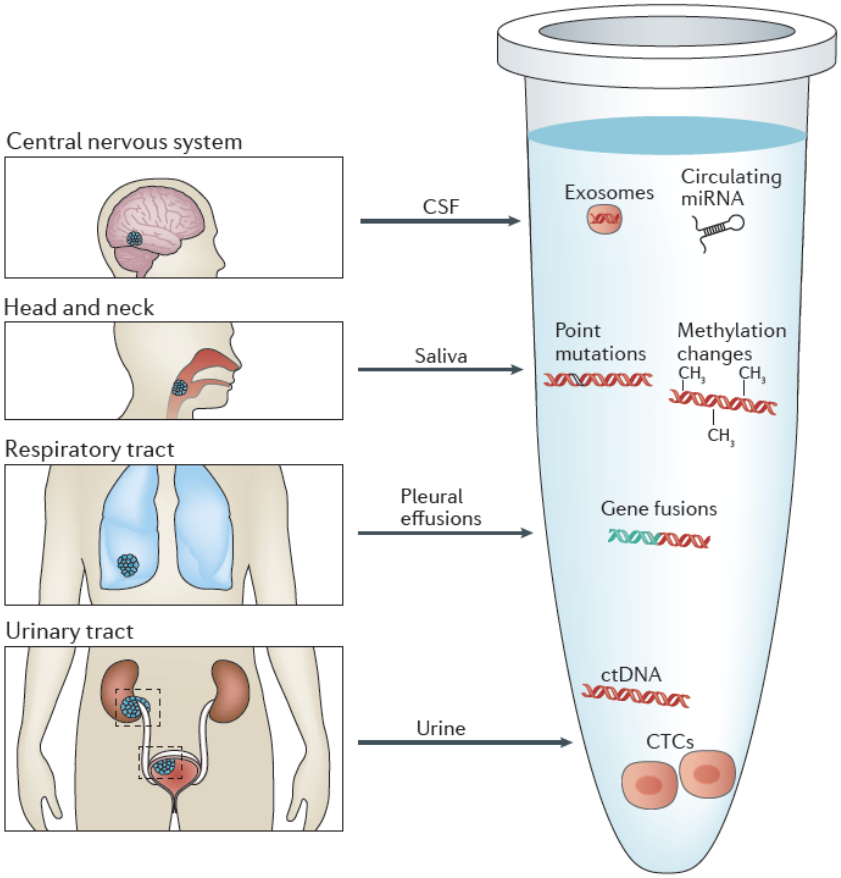


Figure 7: Sources of liquid biopsy. Tumor-derived material can be found in a variety of bodily fluids. Those include exosomes, ctDNA, and circulating tumor cells (CTC). Adapted from Siravegna et al. [30].

1.2.2 Overview of circulating biomarkers and their applications in breast cancer

Circulating biomarkers have been increasingly utilized to reveal the information of diseases at the DNA level for genomic aberration, including mutations and copy number variations; at the epigenome level [31] to detect methylation patterns, and at the transcriptome level [32] for gene expression and fusion detection. Circulating biomarkers have also been used to generate information about the proteome [33-35] and the metabolome [36].

Protein biomarkers such as CA 15-3, CA-125, and prostate-specific antigen (PSA) have been applied as tumor protein biomarkers for breast cancer, ovarian cancer, and prostate cancer, respectively, for decades. However, the sensitivity and specificity of these biomarkers are moderate, especially in early-stage patients [37-40].

In recent years, the most common and well-studied circulating biomarker is called cell-free DNA (cfDNA). cfDNA was first discovered in 1948 by Mandel and Metais [41]. In 1977, Leon et al. found that there were higher concentrations of cfDNA in cancer patients than in healthy individuals [42]. Subsequently in the 1990s, a few groups identified tumor-specific mutations and microsatellite alterations in cfDNA of cancer patients [43-46]. cfDNA has also been applied in noninvasive prenatal test (NIPT) since 1997 when Lo et al. first identified fetal cfDNA in pregnant women [47].

cfDNA is the DNA released into circulation from apoptosis or necrosis of tissue, including tumor tissue or circulating tumor cells (CTCs). cfDNA from tumors is called circulating tumor DNA (ctDNA). It has been demonstrated that ctDNA can be detected in many types of cancer, including: breast [48-52], prostate [53], gastric [54], and others [55]. ctDNA can be defined as any alterations on the sequence, such as tumor-specific

genomic mutations, copy number variations, and rearrangement; or any epigenetic aberrations, such as methylation.

Another widely studied circulating biomarker are circulating tumor cells (CTC). CTCs are cells shed into circulation from primary or metastatic tumors. Tools to isolate CTCs from blood apply the strategy of positive selection, negative selection, or both. CellSearch is the first FDA-approved CTC detector, which positively enriches CTCs that express the epithelial cell adhesion molecule (EpCAM) but do not express the blood-cell marker CD45. CellSearch is suitable for enumeration of CTC, but it cannot separate single CTCs for downstream molecular experiments. In addition, it has been shown that not every CTC expresses EpCAM. Other groups have tried to incorporate different markers into the new systems to enhance CTC enrichment and isolate individual CTCs, such as DEPArray. These single CTCs can then be prepared for single-cell whole genome sequencing (WGS) or whole transcriptome sequencing (WTS).

In 2013, Dawson et al. published an analysis to compare the sensitivity of three biomarkers in metastatic breast cancer: ctDNA, CA 15-3, and CTCs. The result showed ctDNA can be detected in 29 of 30 women (97%); CTCs were detected in 26 of 30 women (87%); CA 15-3 were detected in 21 of 27 women (78%). Not only did ctDNA show the best sensitivity of detection among three biomarkers, it can be used to assess treatment response [52].

Cell-free RNA (cfRNA) has piqued the interest of researchers in the last few years. The cfDNA in the circulation is thought to arise from cells undergoing necrosis or apoptosis. However, cfRNA can be released into circulation from living cells as well. Other than free-floating cfRNA, cfRNA can also come from extracellular vesicles, such

as exosomes. Exosomal RNA (exoRNA) has been employed to detect mutations along with ctDNA. The exoRNA has been characterized to contain microRNA, long non-coding RNA (lncRNA), and mRNA [56-59]. Previous studies have shown that the microRNA and lncRNA profile can be used to determine the tissue-of-origin for a tumor [60-64].

1.3 Technologies to detect aberrations of circulating nucleic acids

1.3.1 Next-generation sequencing

Massively parallel sequencing, or next-generation sequencing (NGS), is a powerful tool with a variety of applications in cancer research. As it pertains to liquid biopsy, considering that the ratio of ctDNA to cfDNA, or the allele frequency (AF) of mutations is extremely low in most cases, the approaches of targeted sequencing for a subset of cancer-related genes are more favorable for ctDNA analysis. The reason being that a targeted-gene approach enables very high coverage of genes which gives better statistical confidence to detect lower AF mutations [65, 66]. There are two major targeted sequencing approaches: amplicon-based and hybrid capture enrichment. Amplicon-based method, such as AmpliSeq, performs highly multiplexed PCR simultaneously: hundreds of genes can be amplified at once. The demand on the quantity (as low as 1 ng) and quality of DNA is also low. The hybrid capture enrichment method, such as CAPP-seq, captures a panel of genes prior to amplification. Therefore, it requires higher input DNA, but the sensitivity of detection could be down to 0.1% of AF (Figure 8). Both of the approaches can include molecular barcodes (or unique molecular identifier; UMI) in the assays to reduce the false positive rate. WGS and whole exome sequencing (WES) had been conducted on cfDNA analysis but it is still not feasible for low AF detection.

There are some advantages of applying NGS to liquid biopsy analysis. First, there is no need to know what mutations are present prior to sequencing. Second, the price of sequencing continues to drop with advances in technology. Third, due to sample barcode technology, more samples can be pooled and accommodated into a sequencing run.

Bioinformatic tools are used to separate sequencing reads afterwards according to sample barcodes.

Currently, there are three major NGS platforms (Figure 9) available for liquid biopsy analysis on nucleic acids: Illumina, Ion Torrent, and Oxford Nanopore.

Illumina uses the “sequencing-by-synthesis” technology. Every cycle it incorporates one corresponding florescence-labelled nucleotide according to the template sequence. Illumina offers a variety of different instruments, but at the high-end the read length can be up to 2 x 300 bp (paired-end reads), and the total output could be up to 6000 Gb.

The sequencing technology of Ion Torrent is based on pyrosequencing. Instead of recognizing individual nucleotides, Ion Torrent utilizes a semiconductor to detect one of the byproducts of pyrosequencing: proton. Since there is no florescence-labelled nucleotide, different nucleotides are introduced into sequencer one at a time: dGTP ->dCTP ->dATP -> dTTP. The average read length is about 400 bp, depending on how many homopolymers in the sequence. The total output could be up to 10 Gb.

Oxford Nanopore can perform single molecule sequencing. Its technology would be able to detect the change of current when a DNA molecule passing through a protein-made nanopore. According to different instruments Oxford offers, the read length can be from hundreds upwards to 2 million bases. The total output could be up to 315 Gb. Although the Oxford Nanopore has the attractive feature of single molecule sequencing, which means no PCR errors will be generated, the sequencing error rate is still high. Considering the demand of high accuracy in the field of liquid biopsy, Oxford Nanopore had been applied in a limited fashion.

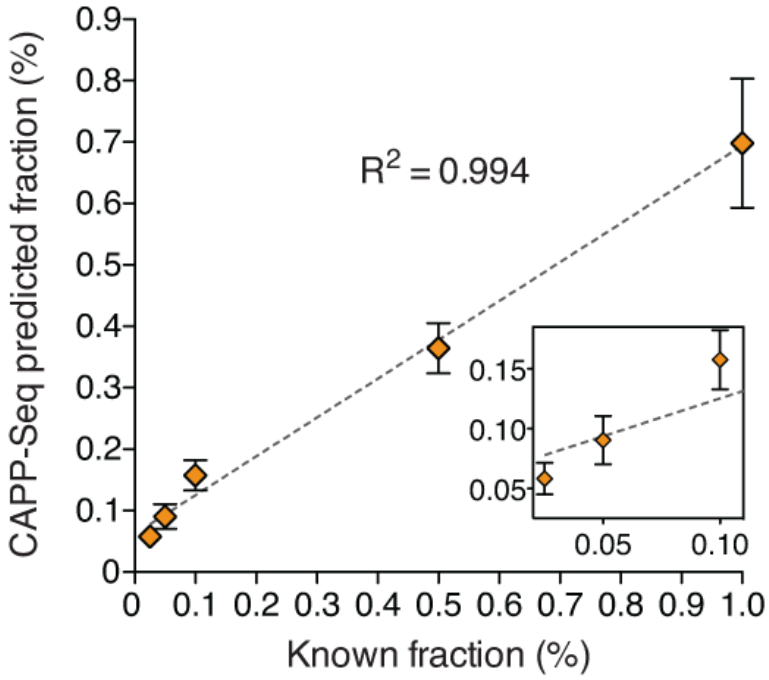


Figure 8: The dilution series analysis of CAPP-Seq. The author spiked in fragmented HCC78 DNA into control cfDNA. Adapted from Newman et al. [67].

Figure 9: Three major next-generation sequencing platforms applied in liquid biopsy. Illumina HiSeq 4000, Ion Torrent, and Oxford Nanopore PromethION. Adapted from Illumina, Thermo Fisher Scientific, and Oxford Nanopore websites.

1.3.2 Digital PCR

Digital PCR (dPCR) is a technology for quantification of DNA or cDNA molecules. Compared to conventional quantitative PCR (qPCR), dPCR can measure the actual number of molecules since dPCR can generate thousands of discrete reactions by forming water-oil emulsion droplets (Figure 10).

Compared to NGS technology, dPCR has better sensitivity of detection (could be down to 0.01% depending on input material). However, to perform dPCR, the known mutations for the targets are required. Additionally, most of the dPCR reactions can detect only one mutation at a time. A few recent publications showed that multiplex dPCR can be performed in one reaction [68-72]. However, the number of mutations for detection is still limited. On the other hand, NGS technology for liquid biopsy can detect hundreds of genes at a time. Furthermore, NGS can also detect short insertion/deletion and rearrangement crossing the coding regions of the genome.

Although there are some the limitations of dPCR, it is still a favorable method for validation and monitoring of known mutations in liquid biopsies, especially for extremely low allele frequency mutations. A number of studies have shown that using both dPCR and NGS on the same study can complement the needs of sensitivity as well as comprehensive genomic profiling [51, 73].

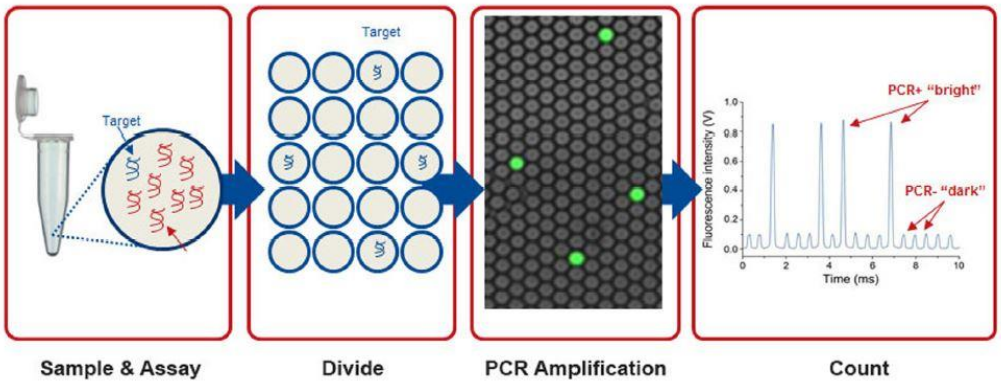


Figure 10: Workflow of digital PCR. Single DNA molecules are separated into individual well. Only the well with target molecules will light up after reactions. Adapted from IDT website.

1.4 Statement of purpose

Standard imaging methods, such as CAT scan and MRI coupled with tissue biopsy, are the gold standard for clinical detection of recurrence. However, these technologies have limited capacity with respect to early detection of minimal residual disease (MRD), which precedes clinical recurrence and usually manifests in distant sites. My long-term goal is to reduce the rate of mortality in patients with chemoresistant TNBC through early detection and therapeutic intervention of patients who harbor MRD. Recently published data has demonstrated that in the metastatic setting, most breast cancer patients have detectable ctDNA in the plasma. I hypothesize that a highly sensitive next-generation sequencing assay using ctDNA and ctRNA to detect somatic mutations will significantly increase the sensitivity to detect MRD. I explored these hypotheses via the following sections:

1. By sequencing and analyzing ctDNA to predict recurrence of TNBC after neoadjuvant chemotherapy, I observed that the overall sensitivity of detection was moderate with high specificity. I was able to detect rapid recurrence but not distant recurrence.
2. By incorporating ctRNA with ctDNA in my sequencing assay, along with more sensitive sequencing methodology, I observed more mutant molecules using my combined ctDNA/ctRNA method, leading to an increased sensitivity to detect MRD.

The following chapters will provide more details of the methods and results of these two aims.

Chapter 2: Circulating tumor DNA detection to predict recurrence of TNBC after neoadjuvant chemotherapy

2.1 Introduction

Triple-negative breast cancer (TNBC) is pathologically defined by the absence of estrogen-receptor (ER), progesterone-receptor (PR), and human epidermal growth factor receptor 2 (HER2) over-expression [2-6]. On the pathological level, TNBC represents only 15-20% of all breast cancers, however, it results in a disproportionally higher rate of mortality [7]. On the molecular level, TNBCs are similar to a molecular subtype called "basal-like breast cancers". Up to 80% of TNBCs are considered basal-like breast cancers [23, 24]. Compared to ER- & HER2-positive breast cancers, TNBCs are more likely to develop brain and visceral metastasis than bone metastasis [10]. They also have a higher likelihood to relapse within the first three years after chemotherapy and surgery, and have a shorter overall survival after the onset of metastatic disease [10]. A significant proportion of patients with TNBC are treated with neoadjuvant chemotherapy. A stark dichotomy exists in outcome based on response to neoadjuvant therapy. Around a third of patients will achieve a pathologic complete response (pCR) and continue to have a favorable overall survival outcome (94% at 3 years). On the contrary, two-thirds of patients have residual disease (RD) after neoadjuvant chemotherapy and are at a high risk of relapse leading to an inferior overall survival (68% at 3 years) [15]. Current standard method to diagnose recurrence of disease is imaging, such as CAT scan and MRI. However, imaging is not an ideal tool to detect MRD. Methods that can trace the presence of tumor-derived material in the circulation of patients who are technically

"disease-free" after neoadjuvant chemotherapy and surgery in the early time point are needed. They may be used to predict those patients whose disease will recur, and further help to determine which patients may need additional post-surgical therapy.

An emerging method for non- or less-invasive cancer diagnosis is the analysis of "liquid biopsies" – the ability to detect tumor characteristics from the circulation, of which the most popular to date, is circulating tumor DNA (ctDNA). ctDNA is secreted into the circulation from the necrosis or apoptosis of tumor tissue, or from CTCs present in blood [55, 74]. It has been shown that ctDNA can be detected in many types of cancer [48-55]. In many cases, the fraction of ctDNA to total cell-free DNA (cfDNA) can be quite small (less than 1%) [75-77]. New technologies such as next-generation sequencing (NGS) and droplet digital PCR (ddPCR) can be used to detect low amounts of ctDNA. Because somatic mutations provide intrinsic specificity for nucleic acid material derived from tumor tissue, the presence of ctDNA implies the presence of disease.

In this chapter, I utilized NGS technology to detect ctDNA from plasma samples of TNBC patients who had neoadjuvant chemotherapy and surgery. By detecting ctDNA, I was able to identify which patients who had MRD prior to clinical detection of recurrence.

2.2 Materials and methods

2.2.1 Clinical trial and correlative samples

BRE09-146 was a prospective, multi-site, randomized Phase II clinical trial of cisplatin +/- PARP inhibition in TNBC patients who have residual disease after neoadjuvant chemotherapy. Eligibility criteria required residual disease, defined as either: (i) residual tumor >2 cm in the breast; (ii) lymph node involvement; or (iii) residual cancer burden (RCB) classification of II or III. Eligible patients were then randomized either to cisplatin for four cycles or cisplatin plus the PARP inhibitor Rucaparib for four cycles followed by maintenance Rucaparib for 24 weeks (Figure 11). Patients were enrolled on trial from March 2010 to May 2013. BRE09-146 is registered on ClinicalTrials.gov (https://clinicaltrials.gov/ct2/show/study/NCT01074970). For correlative analyses, clinical sites submitted tumor biopsies from the time of diagnosis, tumor from residual disease at the time of surgery, as well as whole blood prior to treatment. From the combination arm only (cisplatin + Rucaparib), plasma samples that were originally collected for pharmacokinetic analyses, were obtained at four pre-defined timepoints: Cycles 1 and 2 of the combination phase, and during Weeks 1 and 5 of the maintenance phase (detailed in Figure 11). For this correlative study of ctDNA, each evaluable patient had to have at least one tumor sample (with 60% or greater tumor cellularity), one whole blood sample, and one plasma sample submitted. Tumor DNA was isolated from formalin-fixed paraffin embedded (FFPE) tissue using the Qiagen AllPrep DNA/RNA FFPE kit (Qiagen, Cat. No. 80224). Whole blood was isolated using

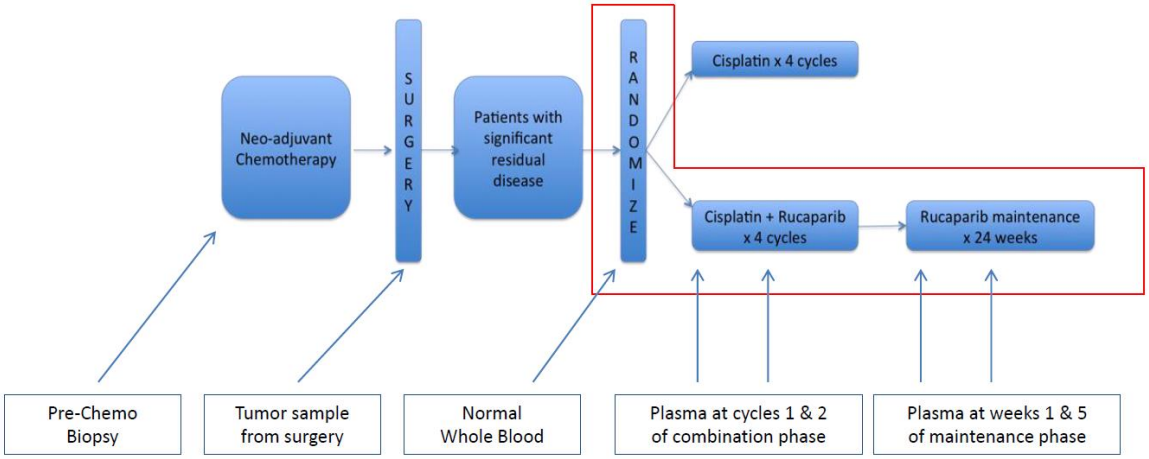


Figure 11: Trial schema for BRE09-146.BRE09-146 was a Phase II clinical trial to evaluate 2-year disease-free survival (DFS) in TNBC patients, treated with either cisplatin (Arm A) or cisplatin in combination with PARP inhibitor Rucaparib (Arm B) after neoadjuvant chemotherapy. Tumor tissue, whole blood, and plasma from four time points after surgery were collected as indicated. In this trial, plasma samples were collected only in Arm B of (the area enclosed by the *red rectangle*). Plasma samples were collected at four timepoints: Cycles 1 and 2 of the combination phase, and during weeks 1 and 5 of the maintenance phase.

AutogenFlex Star instrument and the Flexigene AGF3000 blood kit at the Indiana Clinical and Translational Sciences Institute Specimen Storage Facility (ICTSI-SSF). Plasma DNA was isolated from 1 ml of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen, Cat. No. 55114). All DNA samples were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, Cat. No. Q32851). The trial and correlative studies were approved by the Indiana University Institutional Review Board (IRB); patients provided written informed consent prior to study entry including consent for blood samples for genomic analysis. The study was conducted in accordance with appropriate protocols established by Indiana University.

2.2.2 Library preparation and sequencing

DNA samples from each tumor, blood, and plasma sample underwent the same procedure for library preparation. DNA samples were amplified using a highly-multiplexed polymerase chain reaction (PCR) that amplifies 134 genes that are well-known to be mutated in cancer (Ion Ampliseq Oncomine Research Panel) (See Figure 12 for details). Libraries were completed using the Ion Ampliseq Library Kit. The PCR program for amplification using the Ion Ampliseq Oncomine Research Panel was as follows: 1 x (99°C for 2 minutes), 21 x (99°C for 15 seconds, 60°C for 4 minutes), hold at 10°C. After amplifying DNA targets, DNA amplicons of the same sample were combined in one well of a 96-well plate. The two ends of the amplicons were then partially digested by FuPa Reagent (50°C for 10 minutes, 55°C for 10 minutes, 60°C for 20 minutes, hold at 10°C up to 1 hour), followed by barcoded adapter ligation (22°C for 30 minutes, 72°C at 10 minutes, hold at 10°C up to 1 hour). The libraries were then

purified by AMPure XP reagent at room temperature. The concentration of the eluted library was quantified by qPCR with the Ion Library Quantitation Kit (Cat. No. 4468802) using the Life Technologies 7900HT Fast Real-Time PCR System. The PCR program was as follows: 1 x (95°C for 20 seconds), 40 x (95°C for 1 second, 60°C for 20 seconds). A library concentration of 50 picomolar or greater qualified for subsequent steps of Ion Chem emulsion PCR and templating and sequencing. The libraries were subjected to emulsion PCR, and prepared for sequencing using the Life Technologies Ion Chef and the Ion PI IC 200 Kit (Life Technologies). Up to seven different barcoded libraries were loaded onto one Ion PI v2 BC chip to obtain appropriate coverage. Sequencing was carried out using a Life Technologies Ion Proton Next-generation sequencer (Figure 13). Each sample in my study was sequenced to at least 2500× coverage, with a median coverage of 6071× (range 2559×–13995×). Coverage details and allele frequencies of all observed mutations are detailed in Table 1.

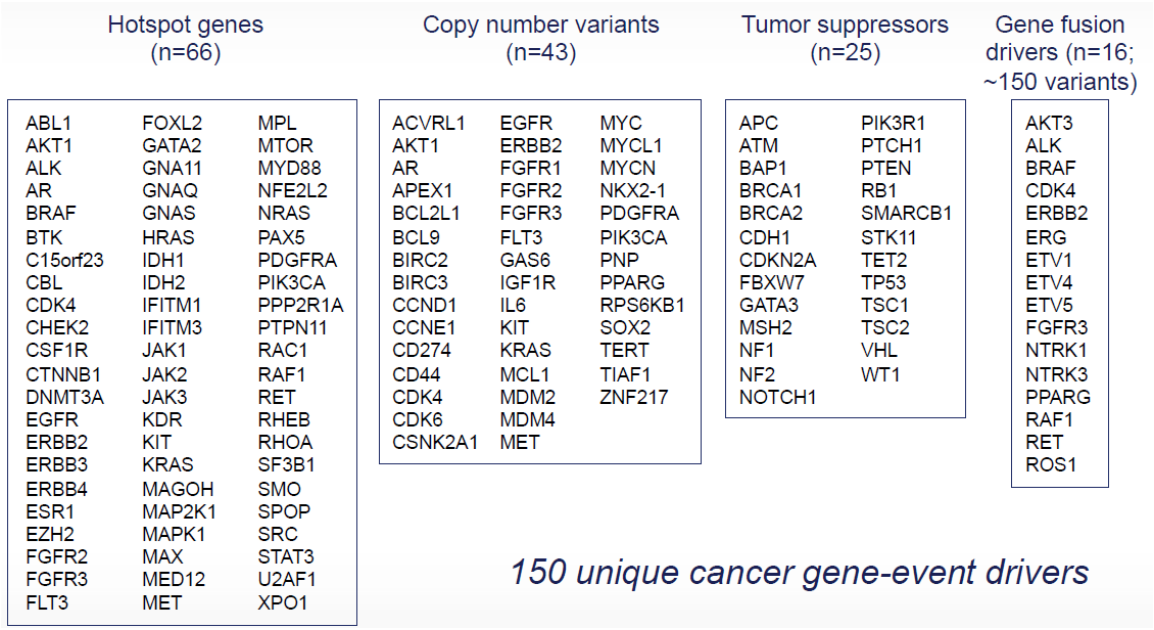


Figure 12: The complete gene list of Ion AmpliSeq Oncomine Research Panel. This panel covers 134 of the most relevant cancer-related genes. Most of the genes are covered the hotspot region; more important tumor suppressor genes are full-length covered. This panel can detect copy number variations on 43 genes and detect gene fusion in 16 genes.

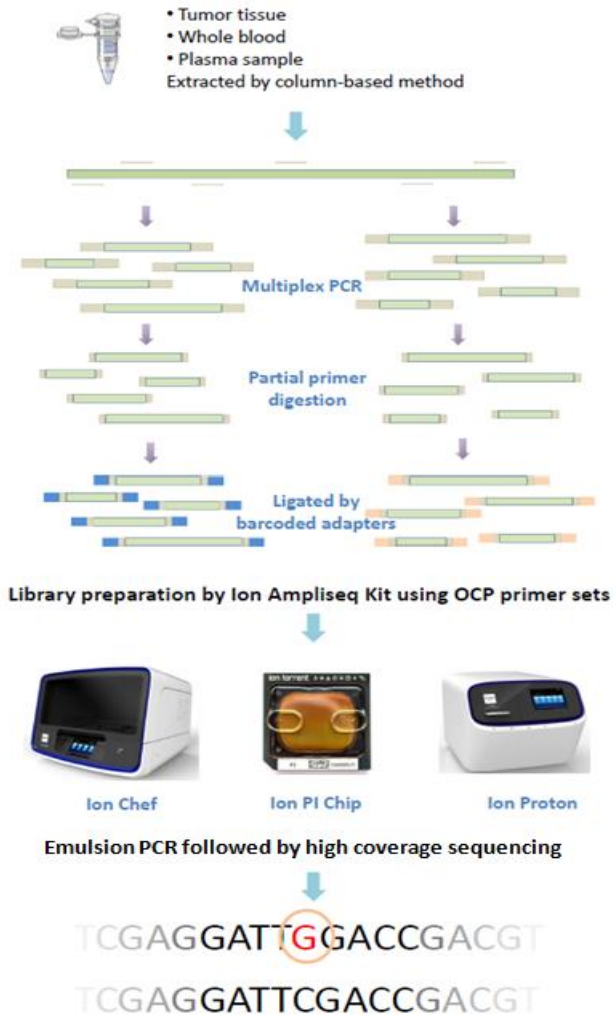


Figure 13: Experimental workflow of mutation identification. DNA from tumor tissue, whole blood and plasma samples was extracted by a column-based method. Isolated DNA was amplified parallelly using the Ion Ampliseq Oncomine Cancer Panel. Amplicons from different samples of the same patient were further ligated by barcoded adapters. Libraries were then processed by an Ion Chef for preparation for sequencing on an Ion Proton next-generation sequencer using the Ion PI chip. Mutations were called using the Torrent Suite v4.2.1 and Torrent Variant Caller v4.2.1.0 software, followed by manual inspection.

ID	Gene	Position	Mutation	Coverage (% mutation allele if applicable)						
				Tumor		Plasma				Blood
				Diag	Surg	Timepoint 1	Timepoint 2	Timepoint 3	Timepoint 4	
146-0004	TP53	Chr17:7577141	C -> T	10057 (13.1%)	N/A	8691	N/A	11372	N/A	14858
146-0005	TP53	Chr17:7578492	C -> T	N/A	6357 (46.1%)	N/A	13493 (2.2%)	12902	8242 (49.4%)	14182
146-0008	TP53	Chr17:7577120	C -> T	N/A	13404 (18.2%)	15581	N/A	N/A	1124	15946
146-0010	TP53	Chr17:7577120	C -> T	5376 (9.0%)	N/A	N/A	N/A	10379	13646	5409
146-0011	PIK3CA	Chr3:178952085	A -> G	N/A	2828 (65.6%)	N/A	N/A	7141	7107	8036
146-0011	TP53	Chr17:7579349	A -> G	N/A	7368 (62.3%)	N/A	N/A	15910	11195	11190
146-0013	CDKN2A	Chr9:21974792	G -> T	556 (38.5%)	618 (39.0%)	2245 (2.0%)	2282	1879	3609 (2.2%)	1392
146-0013	TP53	Chr17:7574003	G -> A	2042 (34.4%)	3313 (27.2%)	3359 (2.7%)	1930	4938	208 (4.8%)	4653
146-0014	TP53	Chr17:7579414	C -> T	N/A	7854 (34.3%)	25802	N/A	16384	N/A	14594
146-0014	ERBB4	Chr2:212812097	T -> C	N/A	892 (9.4%)	325	N/A	529	N/A	1449
146-0024	TP53	Chr17:7574030	G -> -	4415 (39.7%)	N/A	6972	9449	7232	14244	5069
146-0024	PIK3CA	Chr3:178952085	A -> G	4075 (3.5%)	N/A	3122	3155	3686	4018	5241
146-0027	TP53	Chr17:7578212	G -> A	4334 (47.3%)	N/A	6522	N/A	7625	57	6912
146-0027	SMARCB1	Chr22:24129284	G -> T	1804 (11.6%)	N/A	896	N/A	736	0	1701
146-0028	TP53	Chr17:7578394	T -> C	N/A	4778 (15.5%)	N/A	5617	N/A	N/A	11221
146-0031	TP53	Chr17:7578550	G -> A	4780 (45.8%)	3493 (32.1%)	5936	6033	5894	5698	5735
146-0044	PTCH1	Chr9:98241333	G -> A	N/A	4546 (40.7%)	6449	N/A	N/A	5255	7483
146-0044	TP53	Chr17:7577120	C -> T	N/A	12909 (40.5%)	9597	N/A	N/A	9546	12800
146-0044	TP53	Chr17:7578265	A -> G	N/A	16301 (30.3%)	14863	N/A	N/A	9737	15129
146-0048	PIK3CA	Chr3:178947852	C -> T	24371 (64.0%)	N/A	N/A	8193	10534	19733	14339
146-0048	PIK3CA	Chr3:178952085	A -> G	5065 (76.0%)	N/A	N/A	4578	3858	3029	5916
146-0048	RB1	Chr13:49037866	G -> C	174 (30.0%)	N/A	N/A	1251	756	1283	4911
146-0048	TP53	Chr17:7576851	A -> C	1271 (55.0%)	N/A	N/A	1986	1768	4361	2552
146-0055	TP53	Chr17:7577538	C -> T	4128 (60.6%)	6975 (67.5%)	17106	17612	N/A	N/A	17171
146-0063	BIRC2	Chr11:102237807	T -> G	N/A	2466 (14.1%)	2106	2084	2837	996	2564
146-0063	TP53	Chr17:7577105	G -> A	N/A	12694 (35.7%)	2975	14984	17384	17158	13449
146-0064	CDKN2A	Chr9:21971093	C -> T	N/A	4678 (28.0%)	N/A	7440	4844	3565	7164
146-0064	TSC2	Chr16:2129184	G -> A	N/A	3563 (9.3%)	N/A	7007	6245	6260	7532
146-0064	TP53	Chr17:7578212	G -> A	N/A	5919 (46.8%)	N/A	9016	3486	6581	8430
146-0066	NOTCH1	Chr9:139390869	G -> A	N/A	25638 (18.8%)	1385	N/A	14098	N/A	14839
146-0066	TP53	Chr17:7579366	G -> T	N/A	13467 (57.5%)	8516	N/A	12979	N/A	10506
146-0075	TP53	Chr17:7578458	G -> -	4873 (47.0%)	N/A	18083	N/A	11346	13194	11316
146-0075	TP53	Chr17:7578289	C -> T	3032 (3.6%)	N/A	75	N/A	4267	160	5035
146-0075	PIK3CA	Chr3:178927462	G -> A	2234 (3.7%)	N/A	40	N/A	5430	55	9960
146-0082	PIK3CA	Chr3:178952085	A -> G	N/A	2368 (65.6%)	N/A	4524	8399	4830	4832
146-0082	FGFR1	Chr8:38282206	G -> A	N/A	4518 (32.9%)	N/A	8855	17608	10080	9841
146-0082	PTCH1	Chr9:98220584	A -> G	N/A	5290 (22.6%)	N/A	6999	10831	6096	8155
146-0082	GATA3	Chr10:8100889	A -> -	N/A	4517 (32.4%)	N/A	5528	9270	5360	4674
146-0082	CD44	Chr11:35223858	C -> G	N/A	4683 (12.8%)	N/A	6028	11155	6341	8735
146-0082	RB1	Chr13:48922002	A -> T	N/A	971 (27.9%)	N/A	1278	1542	725	3172
146-0082	TP53	Chr17:7577580	T -> C	N/A	5766 (80.9%)	N/A	7670	12230	8557	8202

ID	Gene	Position	Mutation	Coverage (% mutation allele if applicable)						
				Tumor		Plasma				Blood
				Diag	Surg	Timepoint 1	Timepoint 2	Timepoint 3	Timepoint 4	
146-0085	BRAF	Chr7:140453136	A -> T	N/A	4978 (22.0%)	2507	2404	1511	1650	4268
146-0085	AKT1	Chr14:105246551	C -> T	N/A	2432 (36.1%)	2897	3162	3022	4140	2567
146-0085	TP53	Chr17:7576852	C -> T	N/A	3083 (31.7%)	2713	2402	2798	2145	2417
146-0086	TP53	Chr17:7577594	AC -> -	5479 (36.4%)	5047 (37.2%)	9602	7795	N/A	N/A	8977
146-0093	FGFR3	Chr4:1809344	G -> A	2600 (0.0%)	648 (10.3%)	N/A	2682	N/A	N/A	2911
146-0093	TP53	Chr17:7578526	C -> -	6425 (47.9%)	1860 (51.6%)	N/A	13787	N/A	N/A	8914
146-0094	TP53	Chr17:7574018	G -> A	4611 (27.0%)	5972 (0.0%)	N/A	N/A	N/A	12370	7465
146-0102	AKT1	Chr14:105246551	C -> T	1291 (29.4%)	N/A	11377 (0.4%)	N/A	N/A	N/A	8984
146-0103	TP53	Chr17:7577046	C -> A	8008 (41.4%)	N/A	15169	N/A	10291	11629	11425
146-0104	APC	Chr5:112179214	G -> C	N/A	8594 (31.5%)	10063	9472	6611	8687	11268
146-0104	TP53	Chr17:7577558	G -> -	N/A	2403 (52.6%)	6312	6121	9014	6847	6239
146-0107	TP53	Chr17:7578508	C -> T	N/A	10067 (18.1%)	14118	14067	15129	N/A	9347
146-0107	NF1	Chr17:29556965	T -> C	N/A	847 (16.6%)	387	735	951	N/A	1002
146-0112	TP53	Chr17:7578203	C -> T	7691 (60.4%)	9184 (59.7%)	7676 (36.0%)	N/A	N/A	N/A	8079
146-0114	CDK4	Chr12:58142091	G -> T	2095 (28.5%)	N/A	1367	N/A	N/A	N/A	1678
146-0114	TP53	Chr17:7578270	ATGCTGAGGAGGGGCCAGACCTAAGAGCA -> -	1967 (22.1%)	N/A	6566	N/A	N/A	N/A	8183
146-0116	PIK3CA	Chr3:178916948	TTC -> -	2112 (15.9%)	N/A	N/A	674	605	N/A	3163
146-0116	PTEN	Chr10:89685305	T -> G	2042 (38.1%)	N/A	N/A	160	107	N/A	1088
146-0116	NF1	Chr17:29587529	A -> T	5831 (15.5%)	N/A	N/A	1120	231	N/A	7620
146-0129	TP53	Chr17:7578236	A -> T	2762 (58.0%)	N/A	5393	7068	5034	N/A	5259
146-0133	TP53	Chr17:7578263	G -> A	N/A	6810 (12.0%)	10685	499	N/A	5319	14603
146-0135	PIK3R1	Chr5:67591145	TACTTGATGT -> -	N/A	3992 (17.8%)	N/A	4862	3679	2184	7423
146-0135	TP53	Chr17:7578515	T -> A	N/A	6185 (23.0%)	N/A	10196	9035	18749	8014
146-0135	NF1	Chr17:29665714	CCTAAAAGGC -> -	N/A	1295 (9.8%)	N/A	421	409	20	2883

Table 1: The detailed sequencing results of coverage numbers, allele frequency, and tumor cellularity. N/A: not applicable: patients either did not have blood drawn, or sequencing failed at those timepoints.

2.2.3 Bioinformatics analysis

Each sequencing run produced approximately 56–89 million reads. Reads underwent primary analysis using the Torrent Suite v4.2.1, which includes quality control, read trimming, and mapping to the human genome (hg19). Variant calling was performed using the Torrent Variant Caller v4.2.1.0. Somatic mutations were identified by comparing variants observed in the tumor sample that were not present in a matched blood sample. Identified somatic variants were then searched for in the plasma DNA sequencing using the Torrent Variant Caller. I also manually inspected called variants using the integrative genomics viewer [78, 79] to confirm the presence of variants and to rule out false positives.

2.2.4 Statistical analysis

Clinical follow-up data was provided by the trial contract research organization: Hoosier Cancer Research Network. The median follow-up for DFS for the entire trial was 24 months. DFS analysis was performed using Cox regression (IBM SPSS Statistics version 24) and plotted using the Kaplan–Meier method (Graphpad Prism, GraphPad Software, Inc.). In univariate Cox regression analysis, the detection of ctDNA was significantly associated with inferior DFS. In multivariate Cox regression, when adding RCB and number of positive lymph nodes as covariates, the detection of ctDNA was independently associated with DFS (see Results). Age, race, tumor grade, and Eastern Cooperative Oncology Group (ECOG) performance status were not associated with DFS, and not used as covariates.

2.3 Results

2.3.1 Comprehensive genomic profiling of tumor and normal tissue

2.3.1.1 Patient and sample selection

One-hundred thirty-five patients were enrolled on the BRE09-146 clinical trial. Patient characteristics, including: median age, race, neoadjuvant chemotherapy, radiation therapy, median residual lymph node positivity (LN+), and median RCB are detailed in Table 2. All patients received multiple agent neoadjuvant chemotherapy, with the vast majority receiving a combination of anthracycline, cyclophosphamide, and paclitaxel, followed by surgery and radiotherapy (Table 2). TNBC patients who completed neoadjuvant therapy and had significant residual disease were randomized to either cisplatin monotherapy or the combination of cisplastin plus rucaparib. Plasma samples used for the analysis of ctDNA were only collected in patients enrolled in the combination arm (Figure 11). Details of patient selection included in this study are outlined in the Consolidated Standards for Reporting Trials (CONSORT) diagram. In total, 38 patients with matched tumor tissue, blood, and at least one plasma sample were successfully sequenced (Figure 14).

2.3.1.2 Identification of somatic mutations in the primary tumors

I first identified somatic mutations present in the primary tumor by identifying variants from tumor sequencing that were not present in the matched normal blood. Of the 38 patients described above, I successfully identified at least one somatic mutation in

87% of patients (33 of 38; Figure 15), and two or more somatic mutations in 55% of patients (21 of 38; Figure 15). Among those who had somatic mutation(s) identified, 31 patients had TP53 mutations (33 TP53 mutations in total; two patients had dual TP53 mutations). Ten out of 38 patients carried genetic alterations in the genes involved in PI3K signaling pathway. Among those, PIK3CA was the most common gene with genetic alterations (six patients), followed by AKT1 (two patients), PIK3R1, and PTEN (one patient each). The high-rate of TP53 and PI3K mutations is congruent with published data from the The Cancer Genome Atlas (TCGA)[80], which observed the same pattern in TNBC tumors.

2.3.2 Detection of tumor-related mutations in plasma sequencing

I then searched for the somatic mutations identified from the primary tumors in the matched plasma samples. Of the 33 patients who had a somatic mutation identified in their primary tumor, I was able to detect somatic mutations in the plasma of four patients (three TP53 mutations, one AKT1 mutation, and one CDKN2A mutation). All four patients had recurrence of their disease (100% specificity), but sensitivity was limited to detecting only 4 of 13 patients who relapsed (31% sensitivity). Figure 16 to Figure 19 details the time-course of the mutational allele frequency for these four patients. In patient 146-0005 (Figure 16), a TP53 mutation (Chr17:7578492, C to T) was detected in timepoint 2 and 4 plasma samples. A similar pattern (timepoint 1 and 4) was observed in patient 146-0013 who had a different TP53 mutation (Chr17:7574003, G to A) and a CDKN2A mutation (Chr9:21974792) (Figure17). I was also able to detect somatic mutations in plasma samples from the other two patients who had only one timepoint

plasma sample available (146-0102, AKT1 mutation, Chr14:105246551, C to T; 146-0112, TP53 mutation, Chr17:7578203, C to T) (Figure 18 and Figure 19). All mutations were located in exonic regions. Interestingly, all four patients had a rapid recurrence: average of 4.6 months (0.3, 4.0, 5.3, and 8.9 months). The lead time of detection of the mutation in the plasma to clinical recurrence ranged from 0.07 to 8.87 months.

BRE09-146	Arm A cisplatin (*n* = 65)	Arm B cisplatin + rucaparib (*n* = 70)	Subjects from arm B for this study (*n* = 38)
Median age	48 (27–69)	47 (21–75)	47 (21–66)
Race			
a. African American	20.00%	18.60%	13.20%
b. White	75.40%	72.90%	73.70%
c. Asian	1.50%	4.30%	7.90%
d. Others	3.10%	4.30%	5.30%
Neoadjuvant chemotherapy			
a. Anthracycline	89.20%	88.60%	92.10%
b. Cyclophosphamide	95.40%	90.00%	92.10%
c. Taxane	95.40%	92.90%	92.10%
d. Carboplatin	1.50%	10.00%	10.50%
e. Unknown	3.10%	2.90%	0%
Radiation therapy	87.70%	85.70%	84.20%
Median residual lymph node positivity (LN+)	1 (0–15)	1 (0–38)	1 (0–38)
Median residual cancer burden	2.6 (0–5.0)	2.7 (0–5.3)	3.1 (0–5.3)

Table 2: Clinical characteristics of patients enrolled on the BRE09-146 clinical trial.

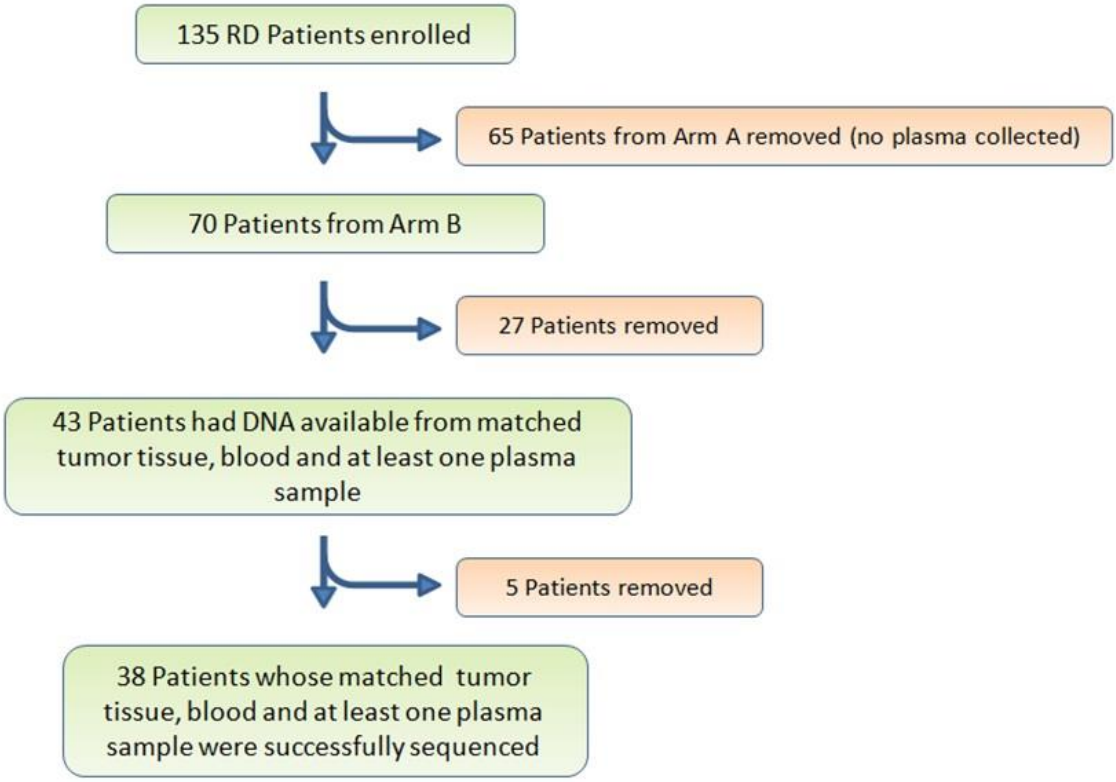


Figure 14: CONSORT diagram. There were 135 patients enrolled in BRE09-146. In this study, I focused on 70 patients from Arm B. In Arm B, 27 patients did not have matched tumor tissue, whole blood, and at least one plasma collection and were excluded from this study leaving an N = 43. A further five patients were removed due to the inability to successfully create a plasma DNA library. In total, 38 patients reached the criteria for analysis.

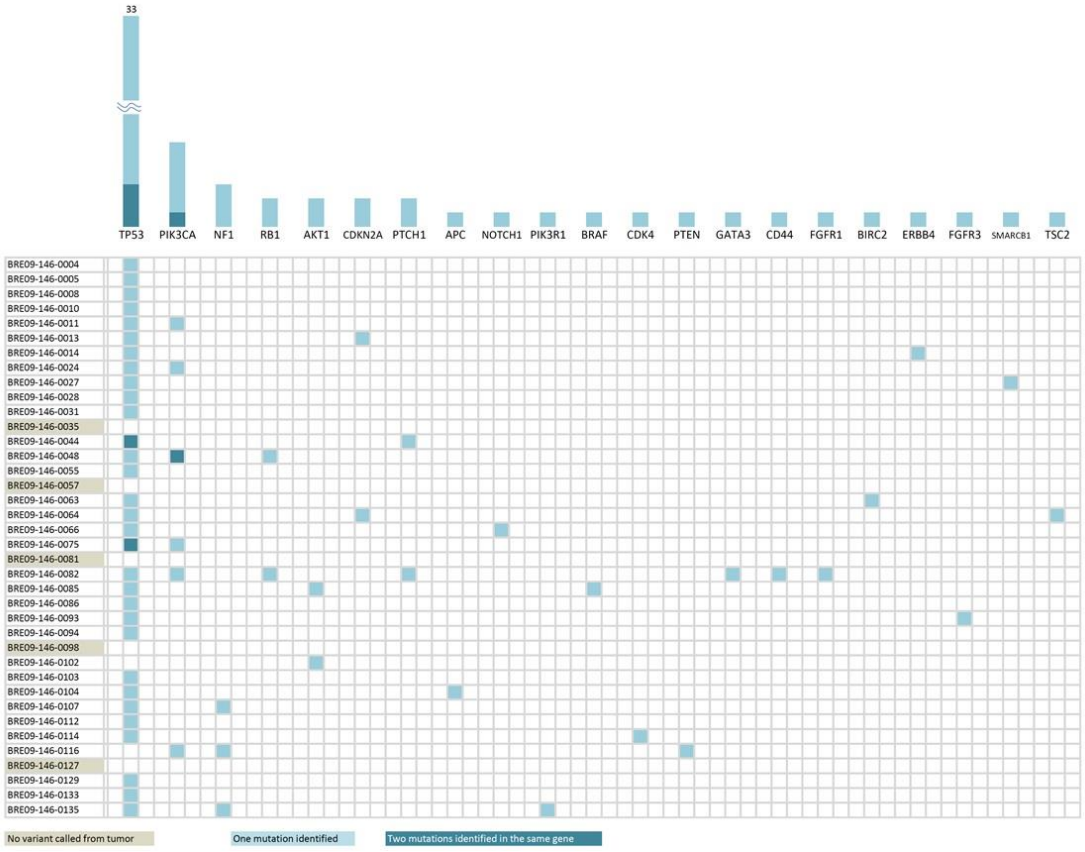


Figure 15: Somatic mutations identified from sequencing of tumor tissues. Among the 38 patients in my study, 33 of them had at least one somatic mutation identified (87%); 21 of them had two or more somatic mutations (55%). TP53 mutations were the most prevalent in this study, followed by PIK3CA pathway mutations. Notably, there were 14 different mutations exclusively present in individual patients, representing the genomic heterogeneity of TNBC patients.

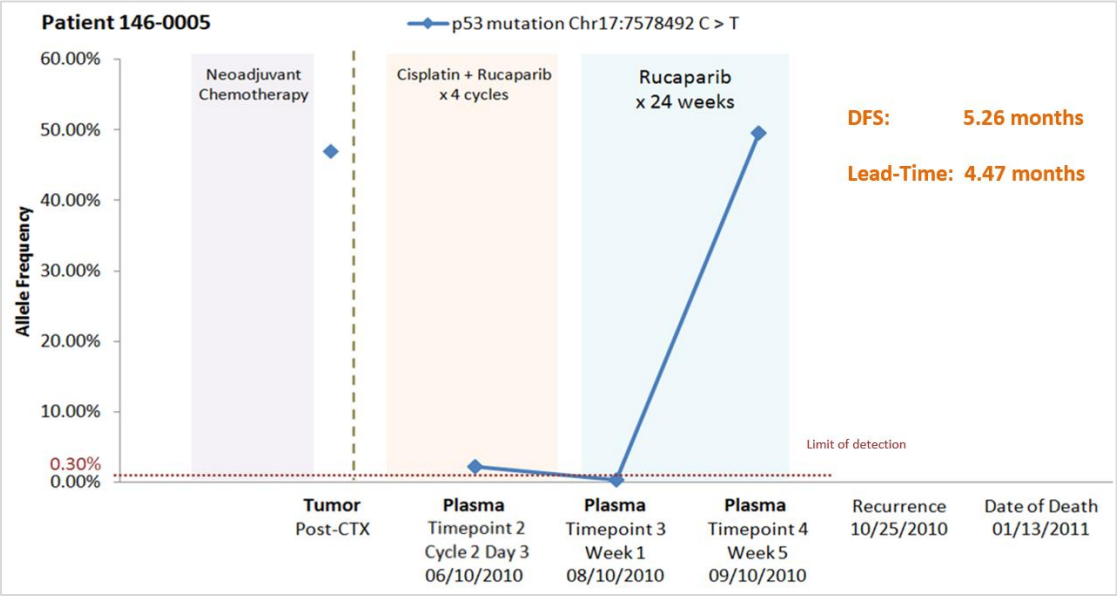


Figure 16: Longitudinal allele frequency tracking of ctDNA mutations in patient 146-0005. The same TP53 mutation was observed at 2 out of 3 timepoints. The increasing allele frequency of ctDNA was observed before clinically recurrence was diagnosed. I was able to detect the ctDNA before clinical recurrence. The lead-time was 4.47 months.

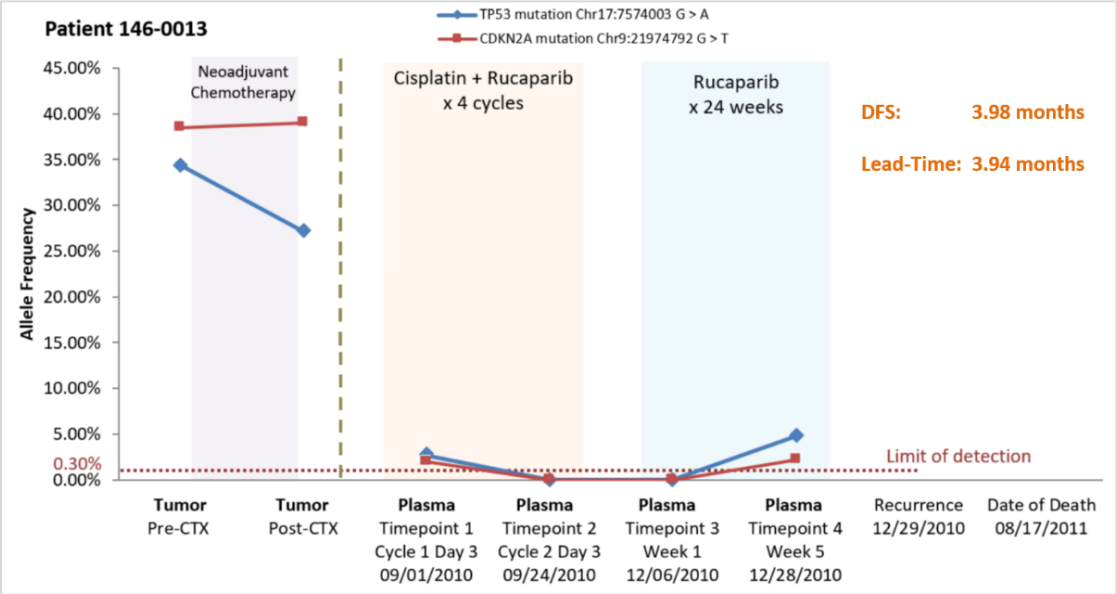


Figure 17: Longitudinal allele frequency tracking of ctDNA mutations in patient 146-0013. Two mutations were detected from different genes: TP53 and CDKN2A. Both mutations were observed at 2 out of 4 timepoints. The increasing allele frequencies of ctDNA were also observed in both genes before clinically recurrence was diagnosed. I was able to detect the ctDNA before clinical recurrence. The lead-time was 3.94 months.

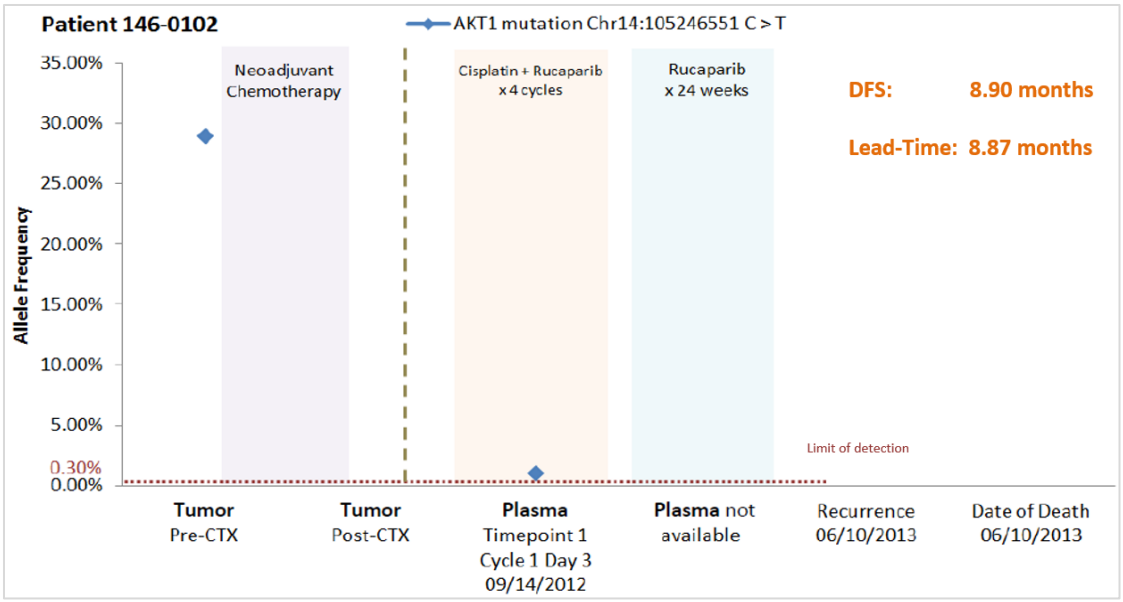


Figure 18: Longitudinal allele frequency tracking of ctDNA mutations in patient 146-0102. AKT1 mutation was observed at the only available timepoint. I was able to detect the ctDNA before clinical recurrence. The lead-time was 8.87 months.

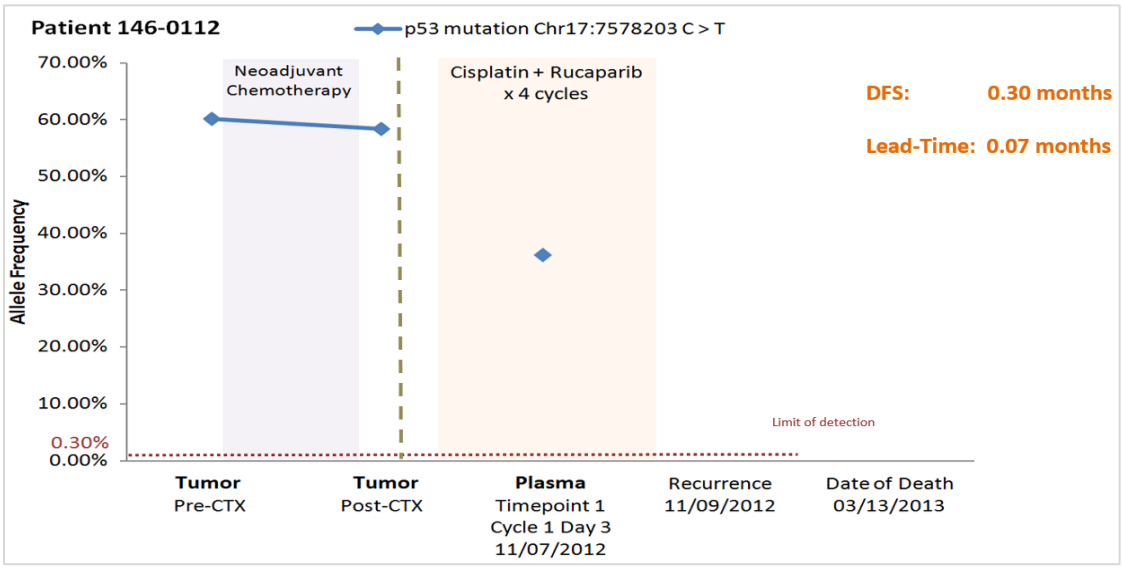


Figure 19: Longitudinal allele frequency tracking of ctDNA mutations in patient 146-0112. TP53 mutation was observed at the only available timepoint. I was able to detect the ctDNA before clinical recurrence. The lead-time was 0.07 months

2.3.3 Correlation of early recurrence and presence of tumor DNA in plasma

A Kaplan–Meier plot demonstrates that the patients who had ctDNA detected in their plasma, had a significantly inferior DFS compared to patients where ctDNA was not detected ($p < 0.0001$, median DFS: 4.6 mos. vs. not reached (NR); hazard ratio (HR) = 12.6, 95% confidence interval (CI): 3.06–52.2) (Figure 20). In multivariate Cox regression analysis, with the addition of RCB and number of positive lymph nodes as covariates, the detection of ctDNA remained independently associated with inferior DFS ($p = 0.011$, median DFS: 4.6 mos. vs. NR; HR = 8.6, 95% CI: 1.6–45.7).

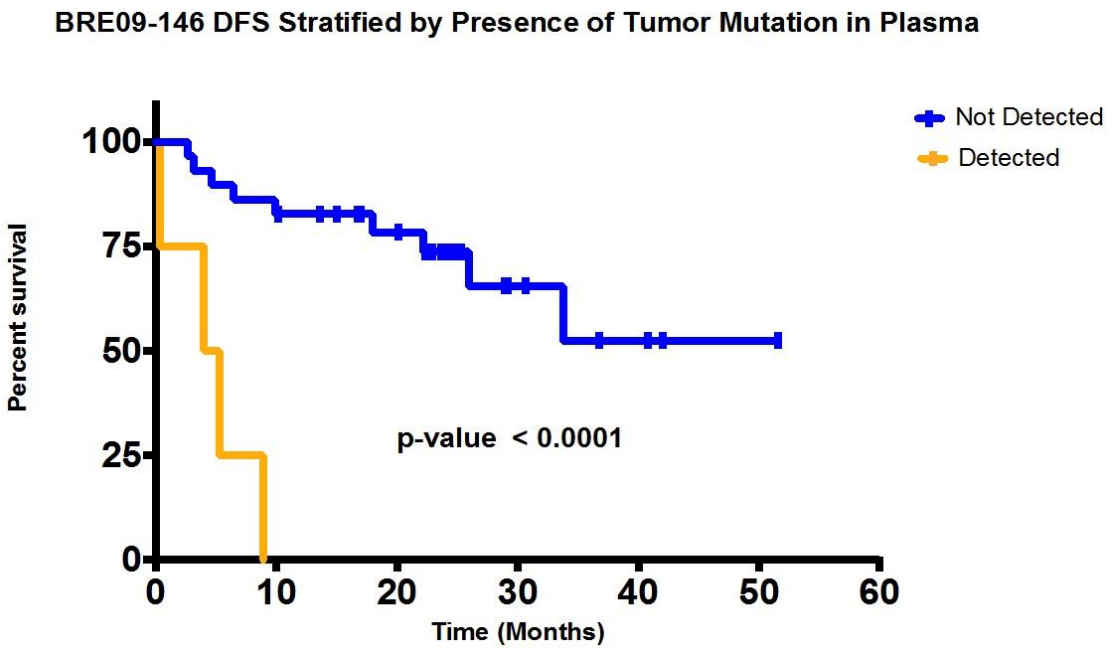


Figure 20: Kaplan–Meier plot: disease-free survival stratified by presence of tumor mutation in plasma. Four patients from this study who had mutation identified from plasma samples relapsed rapidly (0.3, 4.0, 5.3, and 8.9 months). The *yellow line* represents patients with detectable ctDNA in plasma. The *blue line* represents patients with no detectable ctDNA in plasma. The difference in median DFS between patients with detectable ctDNA vs. those without was statistically significant ($p < 0.0001$, median DFS: 4.6 mos. vs. NR; HR = 12.6, 95% CI: 3.06-52.2)

2.4 Discussion

TNBC patients who do not achieve a pCR to neoadjuvant chemotherapy are at a high-risk of recurrence from their disease. Unfortunately, there is no FDA approved standard-of-care for this post-neoadjuvant setting. However, reported results from the CREATE-X trial [81] demonstrated an improvement in 2-year DFS and OS with the use of post-neoadjuvant capecitabine for women with HER2-negative breast cancer with residual disease after neoadjuvant chemotherapy. A subgroup analysis revealed an improved benefit for TNBC patients [81]. Given the clinical scenario, determining those patients who will relapse using methods that can detect the presence of tumor-derived material, even when the patient is technically "disease-free" after surgery, can help predict which patients will recur, and potentially design therapeutic strategies for this population. Although tissue biopsy remains the standard approach for determining the presence of tumor, the liquid biopsy using ctDNA is emerging as a complimentary method. Because somatic mutations provide intrinsic specificity for nucleic acid material derived from tumor tissue, the presence of ctDNA implies the presence of disease. In the evolving realm of circulating biomarkers, a recent study suggests that ctDNA may confer the highest sensitivity. Dawson et al. compared the use of circulating antigen 15-3, CTCs and ctDNA for blood-based detection, and demonstrated that the measurement of ctDNA possessed the highest sensitivity for monitoring metastatic breast cancer [52]. While the vast majority of ctDNA studies have focused on patients with metastatic disease, in this study I focused on patients who are in the curative setting. The patients in my cohort are

disease free by standard clinical assessment, but they are known to be at a high-risk of relapse.

In my study, I searched for somatic mutations in plasma-sequencing data that were first identified in the matched primary tumor. Congruent with published studies of genomic sequencing of TNBCs, I observed a high-rate of TP53 and PI3K pathway mutations [80]. Of 33 evaluable patients, 13 had a clinical relapse, and of those, I was able to detect ctDNA in 4. Of interest, all four of these patients had a rapid recurrence, ranging from 0.3 months to 8.9 months. The lead-time from the first-detection of ctDNA to clinical recurrence ranged from 0.07 months to 8.87 months. I was not able to detect ctDNA in patients with distant recurrence. Further, I was unable to detect ctDNA in five patients who had a recurrence in <12 months. While all patients in which ctDNA was observed did have a rapid recurrence, the low sensitivity to detect distant, and in some cases rapid recurrence, highlights its limitations. Because the ability to detect ctDNA is proportional to the number of mutated molecules in the circulation; disease burden, and the volume of plasma that is sampled are important factors that regulate sensitivity. My study represents a “worst-case scenario” in which there is no detectable disease burden at enrollment, and only 1 mL of plasma in which to perform my studies. Even in this setting, I was able to detect some patients with rapid recurrence. Given the retrospective nature of my study with a limited sample size, however, a prospective trial to prove clinical utility is well warranted.

A pivotal study by Garcia-Murillas et al.in a cohort of early breast cancer patients demonstrated that detection of ctDNA showed a similar pattern of rapid recurrence [51]. A similar study by Olsson et al. showed that serial ctDNA sampling in patients with

primary breast cancer can reach an average lead time of 11 months before the occurrence of metastatic disease [82]. Key differences between my studies and theirs is the specific enrichment of a TNBC population in my study, and my use of NGS vs. ddPCR for ctDNA detection. While ddPCR has increased sensitivity, it requires the generation of patient-specific custom assays, and usually it can only detect one mutation at a time. Thus, NGS has the advantage of being more generalizable for the application of ctDNA detection to a breast cancer population. Lastly, another study by Riva et al., the investigators were unable to detect ctDNA in TNBC patients after surgery using ddPCR [73]. This observation along with mine highlights the importance of serial sampling after surgery.

To summarize this chapter, next-generation ctDNA-sequencing of TNBC patients after neoadjuvant chemotherapy and surgery can detect rapid-recurrence but sensitivity to detect distant recurrence is limited. The following directions can be considered to increase sensitivity of detection: improved or novel extraction methodologies; sequencing chemistries that attempt to provide increased enrichment of mutated DNA molecules (i.e., CAPP-Seq) [67]; sensitive nucleic acid detection using CRISPR-Cas13a along with isothermal amplification [83]; or the combination of ctDNA with other blood-based biomarkers such as cfRNA (mRNA, miRNA, and lncRNA), or exosomal protein [84]. In next chapter, it will be demonstrated that how I incorporated ctRNA into ctDNA detection, along with enhanced extraction methodology and improved chemistries for NGS library preparation.

Chapter 3: Co-detection of circulating tumor DNA and circulating tumor RNA to predict recurrence of TNBC after neoadjuvant chemotherapy

3.1 Introduction

With recent advances in next-generation sequencing technologies, significant progress in ctDNA-based cancer detection has been made. Although multiple studies have demonstrated that the progression of metastatic breast cancer can be tracked using ctDNA [51, 85, 86], few studies have been published demonstrating its use for detection of MRD [51, 87]. Detecting MRD is an evolving and powerful application of this technology. In addition, no studies to my knowledge have been conducted exploring the potential of the combined analysis of ctDNA along with ctRNA. The research proposed is innovative as it leverages cutting-edge and multi-faceted technologies to detect circulating tumor nucleic acids, thus providing unprecedented detection of MRD in early-stage TNBC patients.

As detailed in Chapter 2, using plasma samples from a completed Phase II clinical trial of TNBC patients with residual disease after neoadjuvant chemotherapy, I applied next-generation sequencing to determine if detection of ctDNA can be used as a predictor of relapse in this high-risk patient population. I observed that next-generation ctDNA-sequencing can detect rapid-recurrence but sensitivity to detect distant recurrence is limited. The lack of sensitivity for ctDNA detection can be attributed to both biological and technical factors. In regards to biology, the ability to detect ctDNA is proportional to the number of mutated molecules in the circulation, disease burden, and the volume of

plasma that is sampled. Further, for ctDNA to be present, DNA must be released by dying tumor cells [55]. For patients whose disease does not recur for a long period of time, microscopic deposits of tumor cells can remain in a dormant state and not undergo cellular turnover. RNA however is actively released from live cells in the form of extracellular vesicles known as exosomes. Here I utilize RNA to call mutations. In this chapter I will show that the number of detectable mutated molecules can be increased by simultaneous detection of ctRNA and ctDNA using both ddPCR and NGS. In regards to technical limitations, a particular assay must be able to detect very low concentrations of mutated molecules in the presence of a large concentration of wild-type molecules. Recent advances in library preparation chemistries have now enabled sensitivity to as low as 0.1% allele frequency. For this chapter, I will be employing a newly available chemistry called Avenio (also known as CAPP-seq) which combines a capture-based library preparation along with single molecule barcoding, to achieve high sensitivity [67, 88, 89]. In this chapter, I aim to improve both the biological and technical barriers to achieving higher sensitivity for MRD detection by: (1) incorporation of ctRNA in addition to ctDNA in the same sequencing reaction for simultaneous detection of somatic mutations; (2) applying the novel Avenio method for ctRNA and ctDNA library preparation which uses a capture-based methodology to enrich mutated regions of the genome followed by single-molecule barcoding and sequencing to achieve ultra-low allele frequency detection. I will also show the enhanced cell-free nucleic acid extraction methodology to increase the yield of extraction.

3.2 Materials and methods

3.2.1 ddPCR to assess the contribution of ctRNA to ctDNA on known mutations

Plasma samples were obtained from metastatic cancer patients requested from the IU Simon Cancer Center Tissue Bank. Circulating nucleic acids were isolated from 8 ml of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen, Cat. No. 55114), then split into two groups: (1) ctDNA detection only, (2) ctDNA plus ctRNA detection. In the ctDNA + ctRNA group, circulating nucleic acids underwent reverse transcription using SuperScript IV VILO Master Mix (Thermo Fisher Scientific, Cat. No. 11756050) to obtain cDNA from cfRNA. ddPCR probes and assays were designed to detect the two mutations (TP53 R282W, BioRad Cat. No. 10049550, dHsaMDV2516902; KRAS Q61H, BioRad Cat. No. 10049550, dHsaMDV2010133) that the patient harbored. The ddPCR experiments were performed using QX200 Droplet Digital PCR System.

3.2.2 Clinical samples for ctDNA/ctRNA co-detection using NGS technology

To evaluate the feasibility of co-detecting ctDNA and ctRNA using NGS technology, I conducted the experiment using plasma samples from six metastatic breast cancer patients with known mutations of ctDNA identified previously by Foundation Medicine (FoundationACT). All plasma samples (4 ml of plasma for each patient) were extracted using the Qiagen QIAamp Circulating Nucleic Acids kit (Qiagen, Cat. No. 55114). Each extracted sample was split into two: (1) to be used for detection of ctDNA only; and (2) for detection of ctDNA + ctRNA. Reverse transcription was performed only on the samples in the ctDNA +ctRNA group. SuperScript IV VILO Master Mix (Thermo

Fisher Scientific, Cat. No. 11756050) was used to obtain first-stranded cDNA. Second-stranded cDNA was synthesized by following the "Second-Strand Synthesis" portion of SuperScript™ Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific, Cat. No. 11917010) to obtain double-stranded cDNA. Measurement of DNA and DNA+cDNA were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, Cat. No. Q32851). All samples underwent Avenio ctDNA library preparation.

After confirming that the co-detection of ctDNA and ctRNA can accommodate NGS technology using Avenio capture-based methodology, I then moved on to the additional banked plasma samples from BRE09-146 described in Chapter 2. Plasma DNA was isolated from 1 ml of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen, Cat. No. 55114). All extracted cell-free nucleic acids underwent reverse transcription, second-stranded cDNA synthesis, and quantification as described previously. All samples underwent Avenio ctDNA library preparation.

3.2.3 The Avenio (CAPP-Seq) methodology for ultra-low allele frequency detection

The Avenio technology enables low allele frequency detection of point mutations, indels, fusions, and copy number variation from plasma-derived ctDNA. The Avenio chemistry begins by attaching a unique molecular barcode to each DNA fragment. This assists in downstream bioinformatics analysis to eliminate false positives caused by PCR errors. Subsequent to molecular barcoding, genes of interest are then captured using a liquid hybrid capture (Figure 21). In this chapter, I used the Avenio Expanded Panel which contains probes to hybrid capture 77 cancer genes known to be commonly mutated in cancer (Table 3). This panel includes genes commonly known to be mutated in TNBC,

including: TP53, PIK3CA, PTEN, and RB1 [80]. Single nucleotide point mutations can be detected down to 0.1% allele frequency (Figure 8). Indels and fusions are detectable to 1.0% allele frequency and CNVs down to 2.3 copies (data provided by the manufacturer Roche).

3.2.4 Library Preparation and Sequencing

Extracted cfDNA and reverse transcribed cDNA underwent library preparation using Roche Avenio ctDNA Expanded Kit. DNA fragments were ligated with adapters containing single-molecule molecular barcodes, which enables a reduction in false positives during analysis. This was followed by capture-based target enrichment of 77 cancer-related genes. Samples were sequenced with 2x150 bp paired-end configuration on an Illumina NextSeq 500 using the NextSeq 500/550 High Output kit v2 (300 cycles and 400 million reads per flowcell) at the Indiana University Center for Medical Genomics. Each sample was sequenced to an average of 60 million reads per sample. A phred quality score (Q score) is used to measure the quality of sequencing. A Q30 (99.9% base call accuracy) of 90% or higher in sequencing reads is expected.

3.2.5 Bioinformatics Analysis

Each sample from sequencing had an average of 60 million reads. Reads underwent primary analysis using the Avenio ctDNA Analysis Server which includes quality control, read trimming, demultiplexing, mapping to the human genome (hg19), and variant calling. Variants and allele frequencies are reported by the server and can then be filtered and a report generated. The manual analysis was performed to validate

the expected mutations which were not reported out in the standard pipeline using de-duplicated BAM files on Integrative Genomics Viewer [78, 79].

3.2.6 Statistical Analysis

Clinical follow-up data has already been provided by the trial contract research organization (CRO): Hoosier Cancer Research Network. The median follow-up for DFS for the entire trial is 48 months. I anticipated that the number of patients in which MRD is detected will be higher in the ctRNA+ctDNA samples as compared to ctDNA alone. I also anticipated a proportion of patients will have MRD detected at an earlier timepoint using ctRNA+ctDNA compared to ctDNA alone. I also calculated my sensitivity to predict patients who had a clinical relapse of their disease. The specificity and the DFS analysis will be performed using Cox regression (IBM SPSS Statistics version 24) and plotted using the Kaplan-Meier method (Graphpad Prism, GraphPad Software, Inc.) after conducting additional experiments on samples from patients who did not have a recurrence.

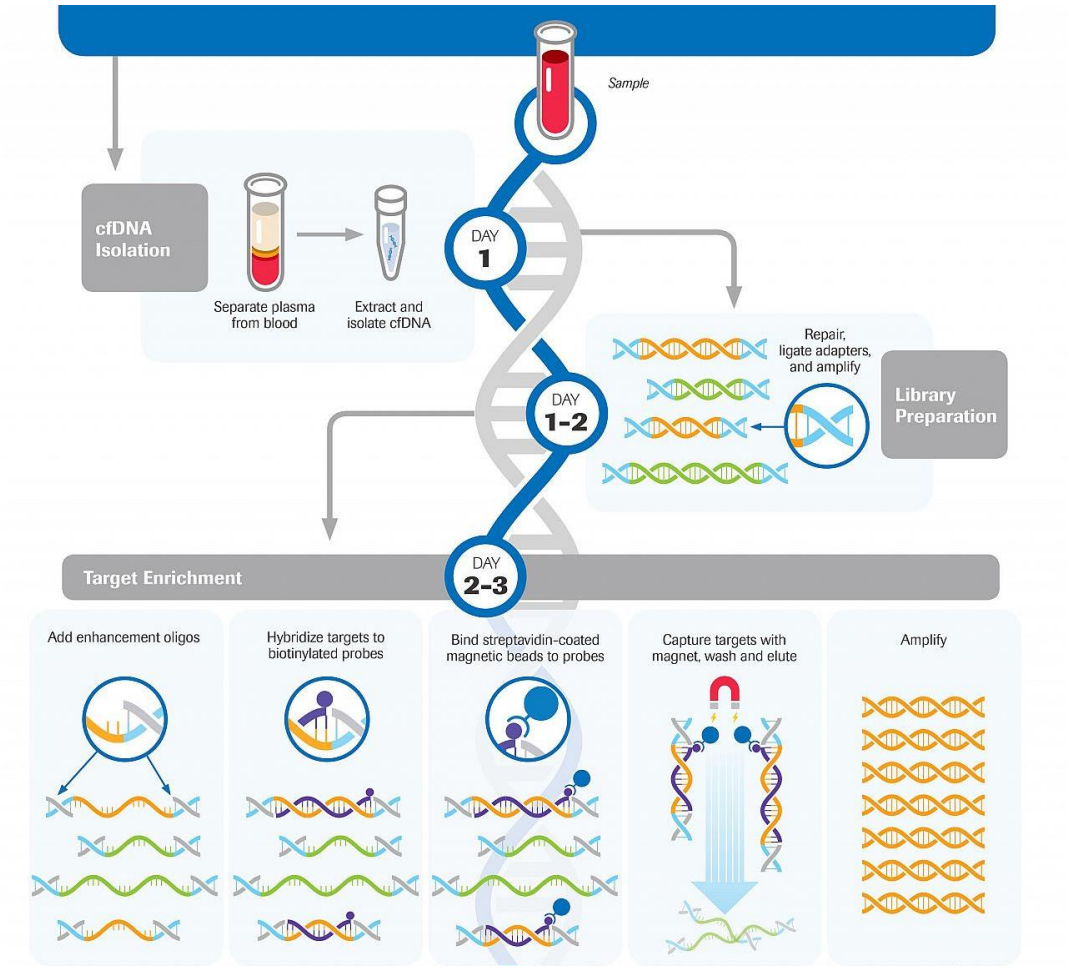


Figure 21: Roche Avenio ctDNA library preparation workflow. cfDNA+cfcDNA or cfDNA-only as input for end repair and adapter ligation, where the molecular barcodes incorporate into each DNA molecules before amplification. During the target enrichment step, 77 genes included in Expanded Kit will be captured and amplified prior to sequencing and bioinformatic analysis.

Assay targets

Gene	Seq Target	SNV	Indel*	Fusion**	CNV**
ABL1	Selected Regions	▪			
AKT1	Selected Regions	▪			
AKT2	Selected Regions	▪			
ALK	Selected Regions	▪	▪	▪	
APC	Selected Regions	▪	▪		
AR	All Coding Regions	▪			
ARAF	Selected Regions	▪			
BRAF	Selected Regions	▪	▪		
BRCA1	All Coding Regions	▪			
BRCA2	All Coding Regions	▪			
CCND1	All Coding Regions	▪			
CCND2	All Coding Regions	▪			
CCND3	All Coding Regions	▪			
CD274	All Coding Regions	▪			
CDK4	All Coding Regions	▪			
CDK6	Selected Regions	▪			
CDKN2A	All Coding Regions	▪			
CSF1R	Selected Regions	▪			
CTNNB1	Selected Regions	▪	▪		
DDR2	Selected Regions	▪			
DPYD	Selected Regions	▪			
EGFR	All Coding Regions	▪	▪		▪
ERBB2	All Coding Regions	▪	▪		▪
ESR1	All Coding Regions	▪			
EZH2	Selected Regions	▪			
FBXW7	All Coding Regions	▪			
FGFR1	Selected Regions	▪			
FGFR2	Selected Regions	▪		▪	
FGFR3	Selected Regions	▪		▪	
FLT1	Selected Regions	▪			
FLT3	Selected Regions	▪			
FLT4	Selected Regions	▪			
GATA3	Selected Regions	▪			
GNA11	Selected Regions	▪			
GNAQ	Selected Regions	▪			
GNAS	Selected Regions	▪			
IDH1	Selected Regions	▪			
IDH2	Selected Regions	▪			
JAK2	Selected Regions	▪			

Gene	Seq Target	SNV	Indel*	Fusion**	CNV**
JAK3	Selected Regions	▪			
KDR	Selected Regions	▪			
KEAP1	All Coding Regions	▪			
KIT	Selected Regions	▪	▪		
KRAS	All Coding Regions	▪			
MAP2K1	Selected Regions	▪			
MAP2K2	Selected Regions	▪			
MET	All Coding Regions	▪	▪		▪
MLH1	All Coding Regions	▪			
MSH2	All Coding Regions	▪			
MSH6	All Coding Regions	▪			
MTOR	Selected Regions	▪			
NF2	All Coding Regions	▪			
NFE2L2	Selected Regions	▪			
NRAS	Selected Regions	▪			
NTRK1	Selected Regions	▪		▪	
PDCD1LG2	All Coding Regions	▪			
PDGFRA	Selected Regions	▪			
PDGFRB	Selected Regions	▪			
PIK3CA	Selected Regions	▪	▪		
PIK3R1	Selected Regions	▪			
PMS2	All Coding Regions	▪			
PTCH1	Selected Regions	▪			
PTEN	All Coding Regions	▪	▪		
RAF1	Selected Regions	▪			
RB1	All Coding Regions	▪			
RET	Selected Regions	▪		▪	
RNF43	Selected Regions	▪			
ROS1	Selected Regions	▪		▪	
SMAD4	All Coding Regions	▪			
SMO	All Coding Regions	▪			
STK11	All Coding Regions	▪			
TP53	All Coding Regions	▪			
TERT Promoter	Selected Regions	▪			
TSC1	Selected Regions	▪	▪		
TSC2	Selected Regions	▪			
UGT1A1***	Selected Regions	▪			
VHL	All Coding Regions	▪			

Table 3: The targets of Roche Avenio ctDNA Expanded Kit. 77 genes are included in this panel to detect point mutations, indels, fusions, and copy number variations. The panel size is around 192 kb which allows to have higher sequencing depth per sample.

3.3 Results

3.3.1 Proof of concept study in metastatic cancer

3.3.1.1 ddPCR

I recently generated preliminary data using digital droplet (ddPCR) on a plasma sample from a patient with metastatic cancer. This particular patient harbored a known TP53 R282W mutation and a KRAS Q61H mutation. The ddPCR probes used to detect both mutations were designed to be contained within a single exon, and did not cross splice junctions. As can be seen in Figure 22, the combination of ctRNA along with ctDNA resulted in a significant increase in the absolute copy number of mutated molecules (TP53 R282W = 47.9% increase; KRAS Q61H = 83.9% increase).

To corroborate my observations, data presented at the 2015 American Society of Clinical Oncology annual meeting by Exosome Diagnostics demonstrated that by using a proprietary column-based extraction strategy (known as EXO52), cell-free DNA as well as exosomal RNA can be simultaneously captured. When analyzed, the addition of reverse-transcribed exosomal RNA to cell free DNA significantly increases the number of gene copies that can be profiled from plasma samples [90] (Figure 23a). Further data from a similar study demonstrated that the increase in copy number of mutated molecules by combining ctRNA with ctDNA can consistently occur at multiple timepoints from the same patient. As seen in Figure 23b, a significant Increase in copy number of mutated BRAF is observed over a period of 6 months when ctRNA is added to ctDNA [91].

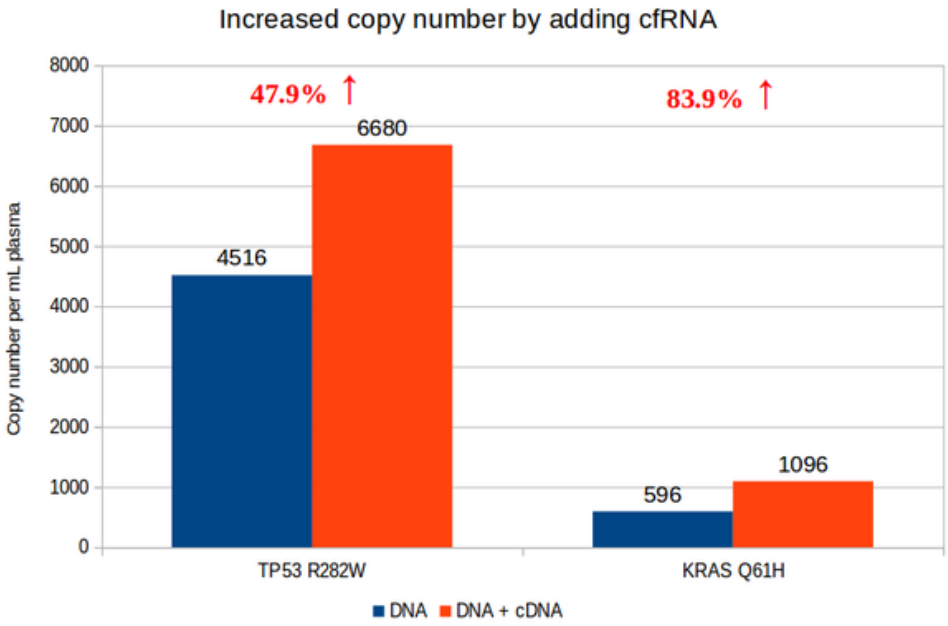
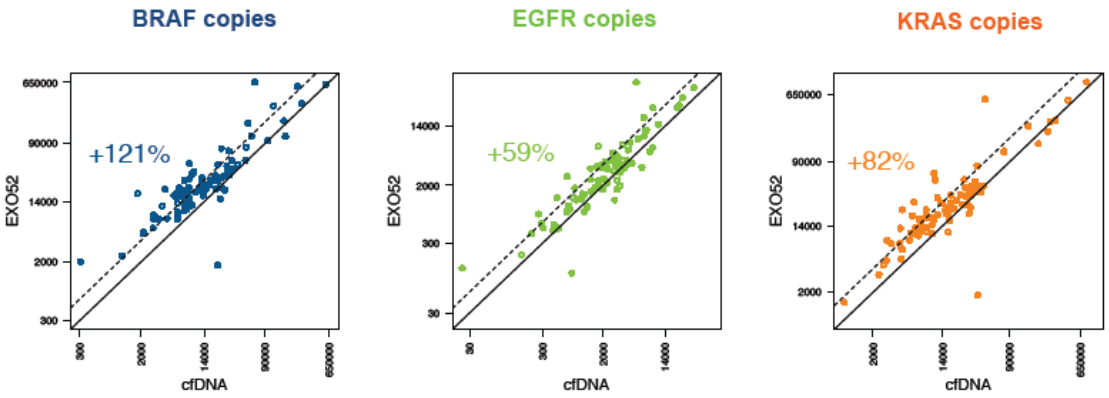


Figure 22: The ddPCR experiment on DNA or DNA + cDNA. Copy number of mutated TP53 and KRAS are increased significantly by adding ctRNA to ctDNA (47.9% and 83.9%, respectively).

a



b

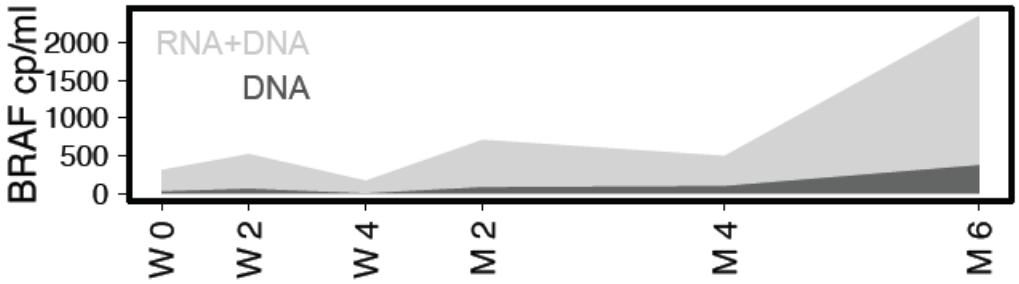


Figure 23: ddPCR to detect ctRNA from other studies. (a) Absolute quantification of extracted gene copies. Comparison of two fractions isolated from 86 different patient samples: cell-free DNA (cfDNA) and cfRNA+cfDNA (EXO52). In these three genes, the added molecules from RNA are around 100% (dotted line). (b) An example from serial blood sampling of a melanoma patient being treated with ipilimumab. Y-axis represents concentration of mutated copies of BRAF per milliliter of plasma. X-axis represents treatment timepoints. As can be seen across timepoints, ctRNA is additive to ctDNA by increasing the concentration of evaluable mutated molecules. Adapted from Enderle et al. [90, 91].

3.3.1.2 Comparison of ctDNA-only and ctDNA/ctRNA co-detection on metastatic breast cancer samples using NGS technology

Plasma samples from six metastatic breast cancer patients were previously sequenced by Foundation Medicine (FoundationACT) to detect ctDNA. I used additional banked plasma samples for this proof of concept study. I was able to identify at least one mutation in both ctDNA-only and ctDNA/ctRNA groups from all six patients. The mutations identified here were concordant to those previously reported by FoundationACT. Five patients had one mutation detected: (1) 0534-417, TP53 p.Arg248Gln; (2) 0534-419: PIK3CA p.His1047Arg; (3) 0534-427: PIK3CA p.His1047Arg; (4) 0534-435: TP53 p.Arg175His; (5) 0534-469: ERBB2 insertion (A775_G776insYVMA). Patient 0534-430 had seven TP53 mutations detected (p.Arg306*, p.Arg273Cys, p.Gly245Ser, p.Arg213*, p.Pro152Leu, p.Arg110Cys, and p.Gly105Ser).

The numbers of mutant molecules in the ctDNA/ctRNA co-detection group were all higher than the corresponding mutations in ctDNA-only group. The average percentage increase was 67% (range=2.8%-385%). The allele frequencies among all mutations were similar to the samples previously sequenced by Foundation Medicine (FoundationACT) (Figure 24-29; Table 4). The copy number of mutation was not available in Foundation Medicine's reports.

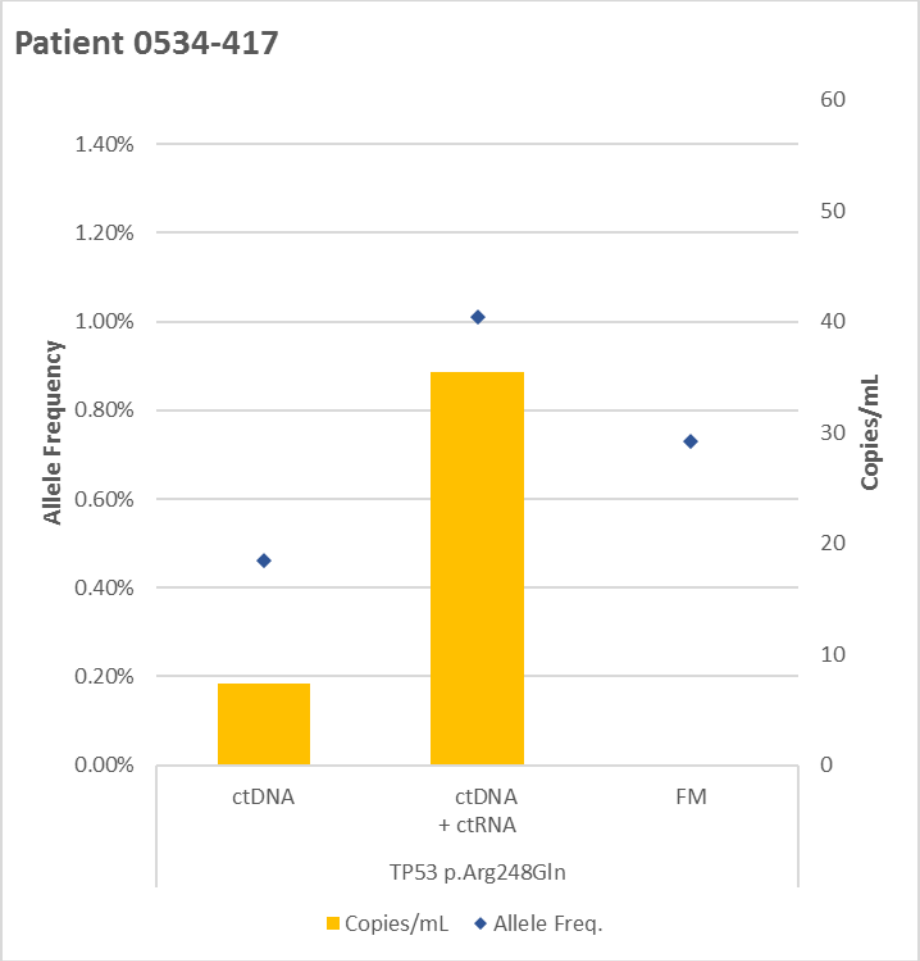


Figure 24: Comparing the sequencing result of patient 0534-417. More mutant molecules (TP53) were detected in ctDNA + ctRNA group per ml of plasma. Previous sequencing result from Foundation Medicine (FM) confirmed the same mutation with similar allele frequency. Copy number per ml from FM was not available.

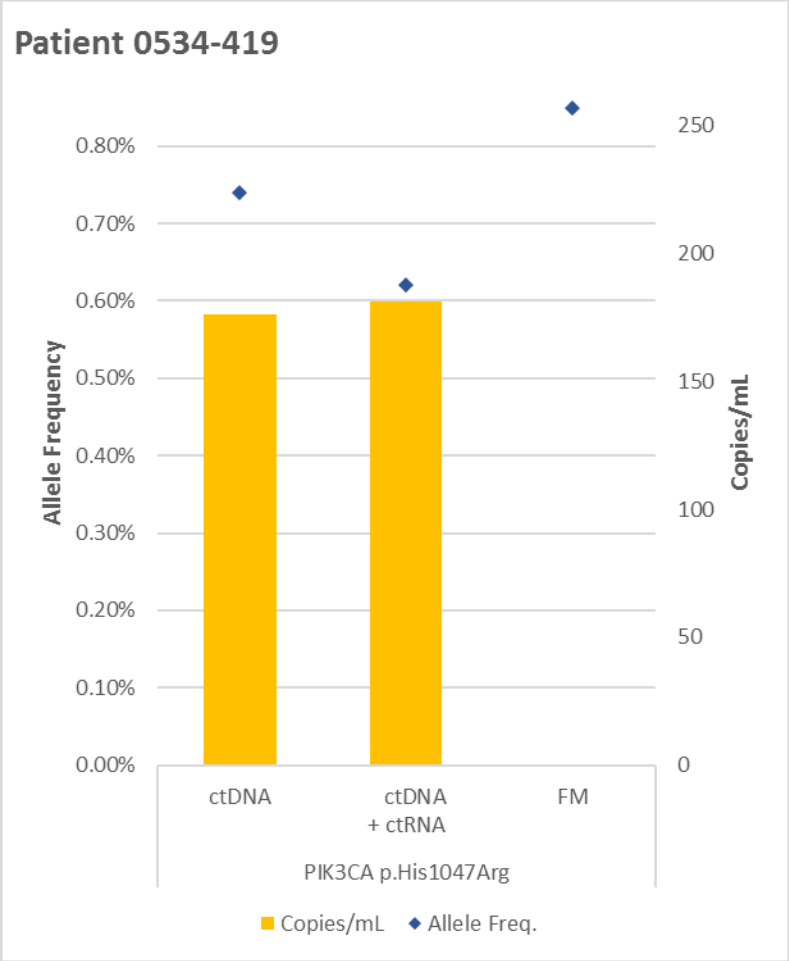


Figure 25: Comparing the sequencing result of patient 0534-419. More mutant molecules (PIK3CA) were detected in ctDNA + ctRNA group per ml of plasma. Previous sequencing result from FM confirmed the same mutation with similar allele frequency. Copy number per ml from FM was not available.

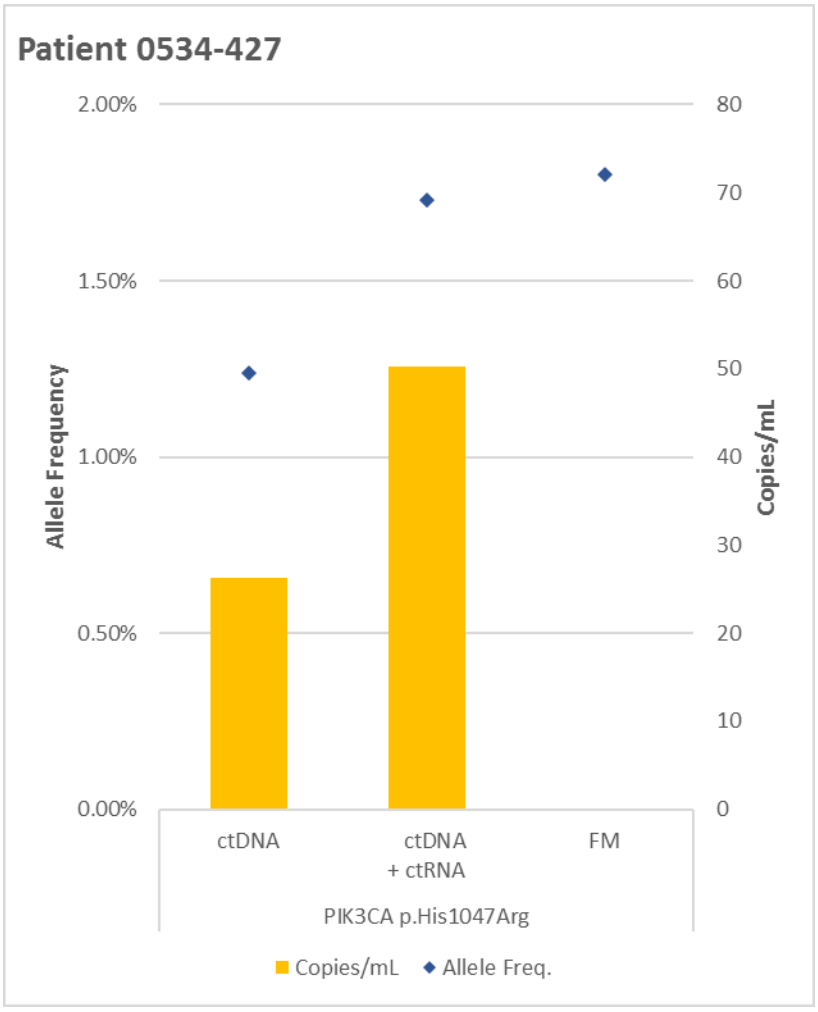


Figure 26: Comparing the sequencing result of patient 0534-427. More mutant molecules (PIK3CA) were detected in ctDNA + ctRNA group per ml of plasma. Previous sequencing result from FM confirmed the same mutation with similar allele frequency. Copy number per ml from FM was not available.

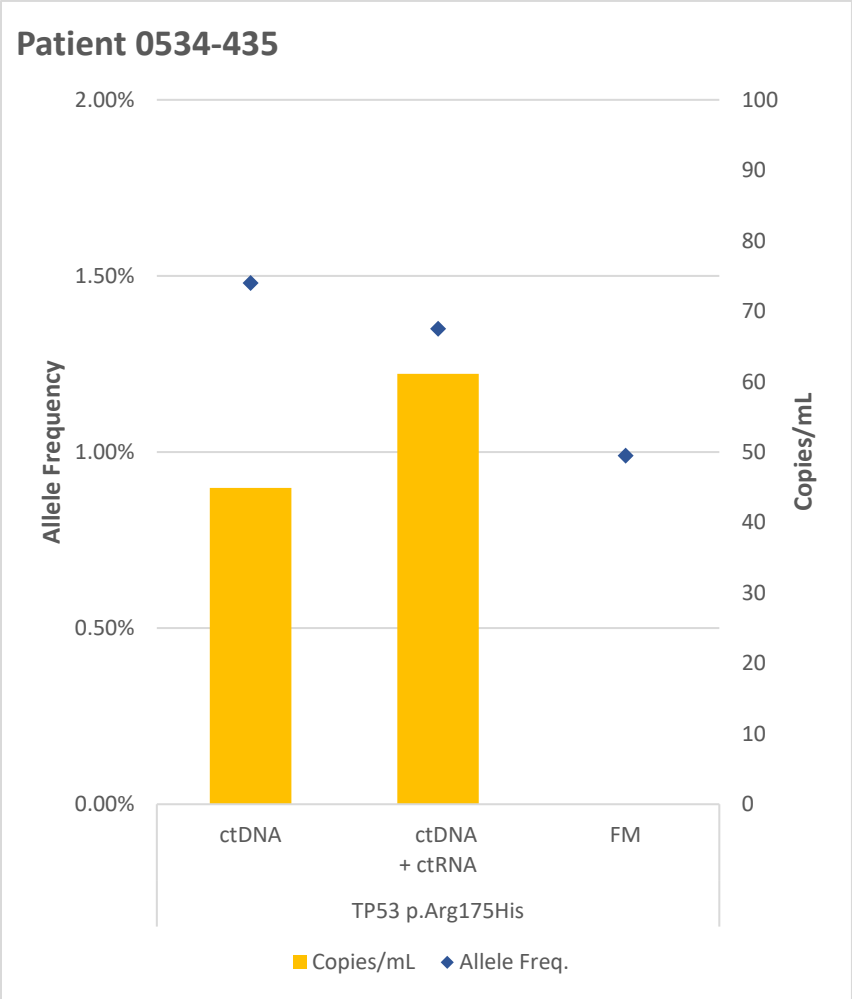


Figure 27: Comparing the sequencing result of patient 0534-435. More mutant molecules (TP53) were detected in ctDNA + ctRNA group per ml of plasma. Previous sequencing result from FM confirmed the same mutation with similar allele frequency. Copy number per ml from FM was not available.

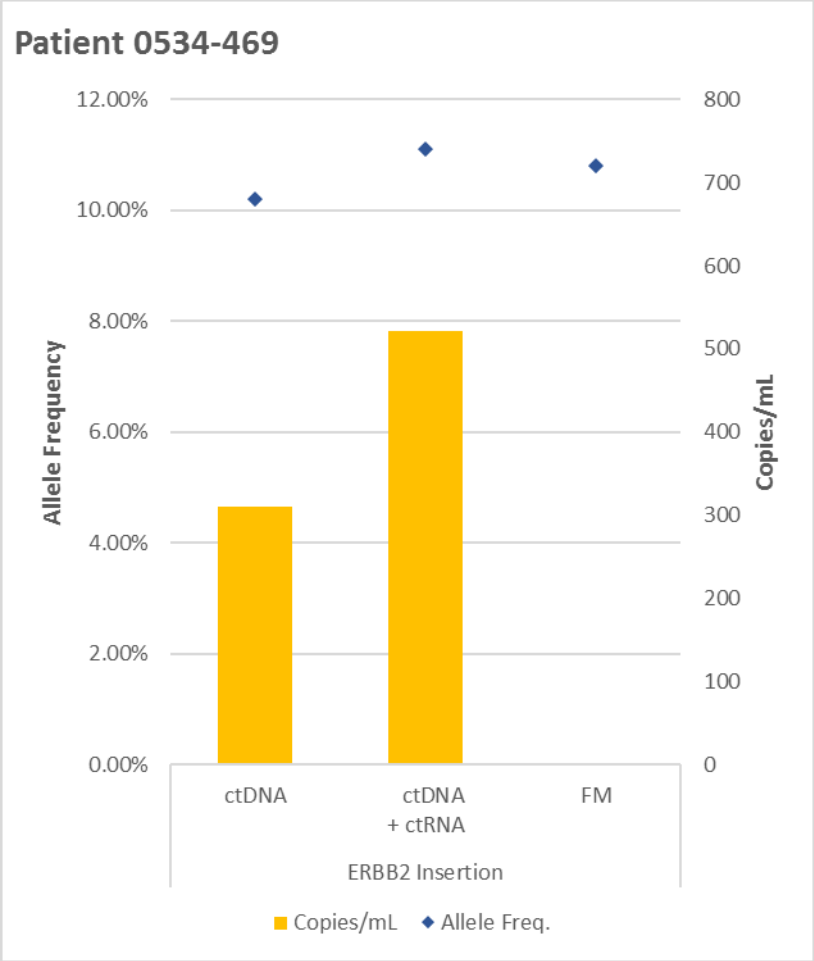


Figure 28: Comparing the sequencing result of patient 0534-469. More mutant molecules (ERBB2) were detected in ctDNA + ctRNA group per ml of plasma. Previous sequencing result from FM confirmed the same mutation with similar allele frequency. Copy number per ml from FM was not available.

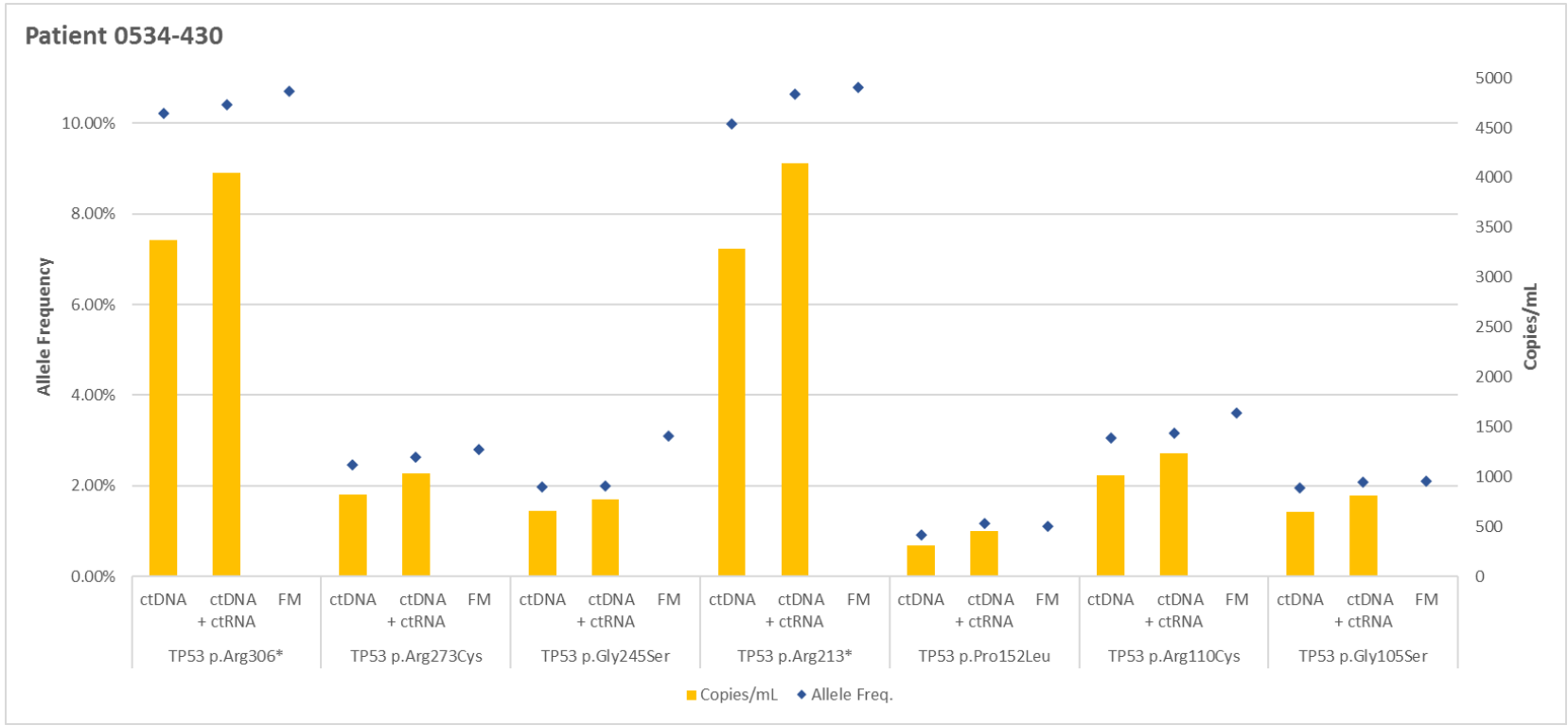


Figure 29: Comparing the sequencing result of patient 0534-430. More mutant molecules (TP53) were detected in ctDNA + ctRNA group per ml of plasma among all seven TP53 mutations. Previous sequencing result from FM confirmed the same mutations with similar allele frequencies. Copy number per ml from FM was not available.

Patient	Gene	Mutation	Allele Frequency			Copies per mL		
			ctDNA	ctDNA/ctRNA	FM	ctDNA	ctDNA/ctRNA	Increased %
0534-417	TP53	p.Arg248Gln	0.46%	1.01%	0.73%	7.3	35.4	384.93%
0534-419	PIK3CA	p.His1047Arg	0.74%	0.62%	0.85%	176	181	2.84%
0534-427	PIK3CA	p.His1047Arg	1.24%	1.73%	1.80%	26.2	50.3	91.98%
0534-435	TP53	p.Arg175His	1.48%	1.35%	0.99%	44.9	61.1	36.08%
0534-469	ERBB2	ERBB2 insertion	10.20%	11.10%	10.80%	310	521	68.06%
0534-430	TP53	p.Arg306*	10.21%	10.41%	10.70%	3370	4050	20.18%
		p.Arg273Cys	2.47%	2.64%	2.80%	816	1030	26.23%
		p.Gly245Ser	1.98%	1.99%	3.10%	652	775	18.87%
		p.Arg213*	9.98%	10.64%	10.80%	3290	4140	25.84%
		p.Pro152Leu	0.92%	1.17%	1.10%	305	455	49.18%
		p.Arg110Cys	3.06%	3.17%	3.60%	1010	1230	21.78%
		p.Gly105Ser	1.95%	2.07%	2.10%	643	807	25.51%

Table 4: Patient-specific mutations and corresponding allele frequencies and mutant copies per ml plasma. Allele frequencies were similar to each other of the same mutation. The numbers of copies per ml of plasma were all higher in the ctDNA/ctRNA group.

3.3.2 Co-detection of mutated molecules in plasma sequencing of early TNBC

I then performed the co-detection strategy on additional banked plasma samples from BRE09-146. To assess if the sensitivity of detection would increase, I focused on plasma samples from 13 patients who had a recurrence. In total, forty-three plasma samples were sequenced. Two of those forty-three samples failed due to insufficient input for library preparation.

In the reports generated from standard pipeline of Avenio ctDNA Analysis Server, I was able to detect somatic mutations in the plasma of five patients (four TP53 mutations, one ATK1 mutation, and one CDKN2A mutation). Figure 30 to Figure 34 detail the time-course of mutational allele frequency and copy of mutation per ml of plasma for these five patients. In patient 146-0005 (Figure 30), a TP53 mutation (Chr17:7578492, C to T) was detected in all four timepoints. A similar pattern was observed in patient 146-0013 who had a different TP53 mutation (Chr17:7574003, G to A; all three timepoints) and a CDKN2A mutation (Chr9:21974792, G to T; timepoint 1 and 3) (Figure31); in patient 146-0102 who had an AKT1 mutation (Chr14:105246551, C to T) in three of four timepoints (Figure 32); in patient 146-0135 who had a TP53 mutation (Chr17:7578515, T to A) in two of four timepoints (Figure 33). I was also able to detect somatic mutations in plasma sample from one patient who had only one timepoint plasma sample available (146-0112, TP53 mutation, Chr17:7578203, C to T) (Figure 34). All five patients had a rapid recurrence: average of 4.98 months (0.3, 4.0, 5.3, 6.4, and 8.9 months). The lead time of detection of the mutation in the plasma to clinical recurrence ranged from 0.07 to 8.87 months.

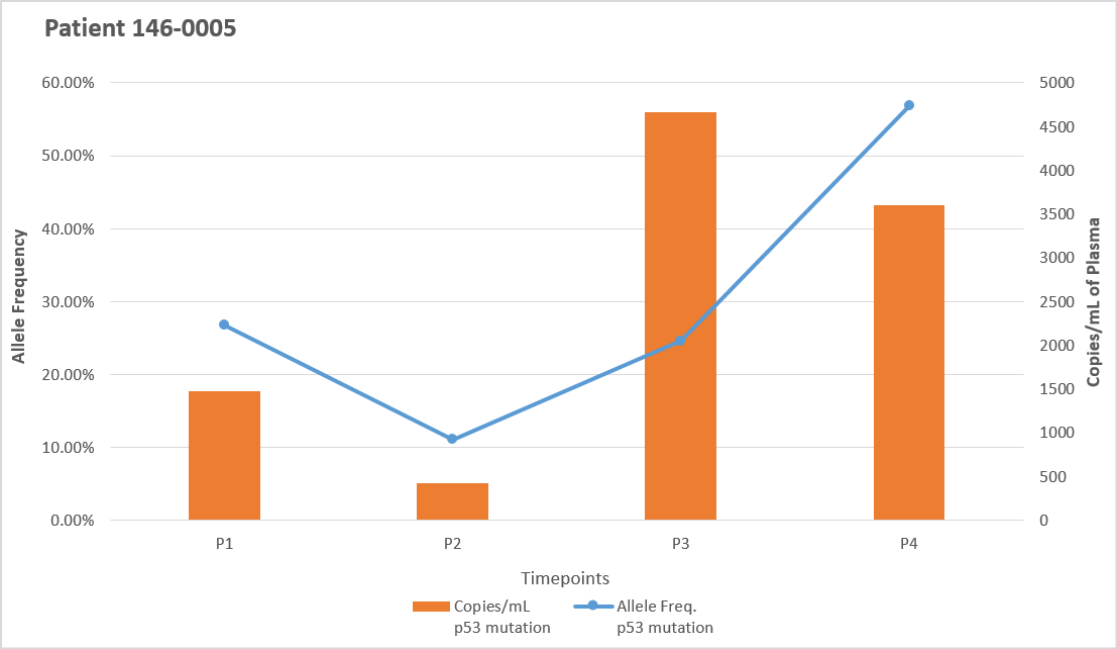


Figure 30: Longitudinal allele frequency tracking of ctDNA/ctRNA mutations in patient 146-0005. The same TP53 mutation was observed at all four timepoints.

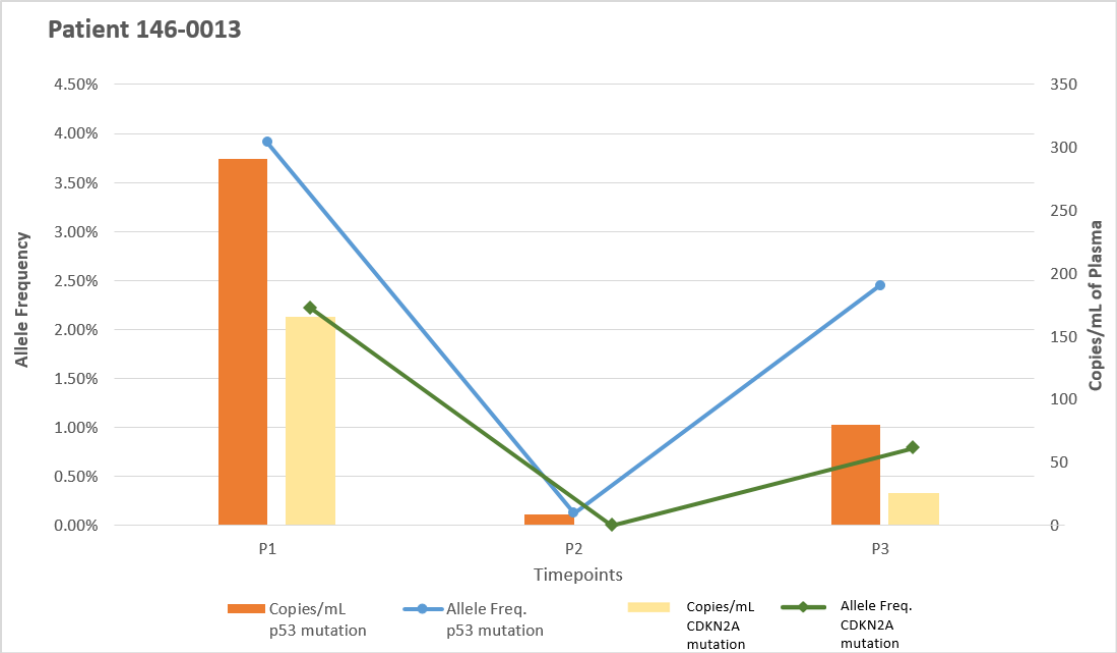


Figure 31: Longitudinal allele frequency tracking of ctDNA/ctRNA mutations in patient 146-0013. The same TP53 mutation was observed at all three timepoints. CDKN2A mutation was observed at timepoint 1 and timepoint 3.

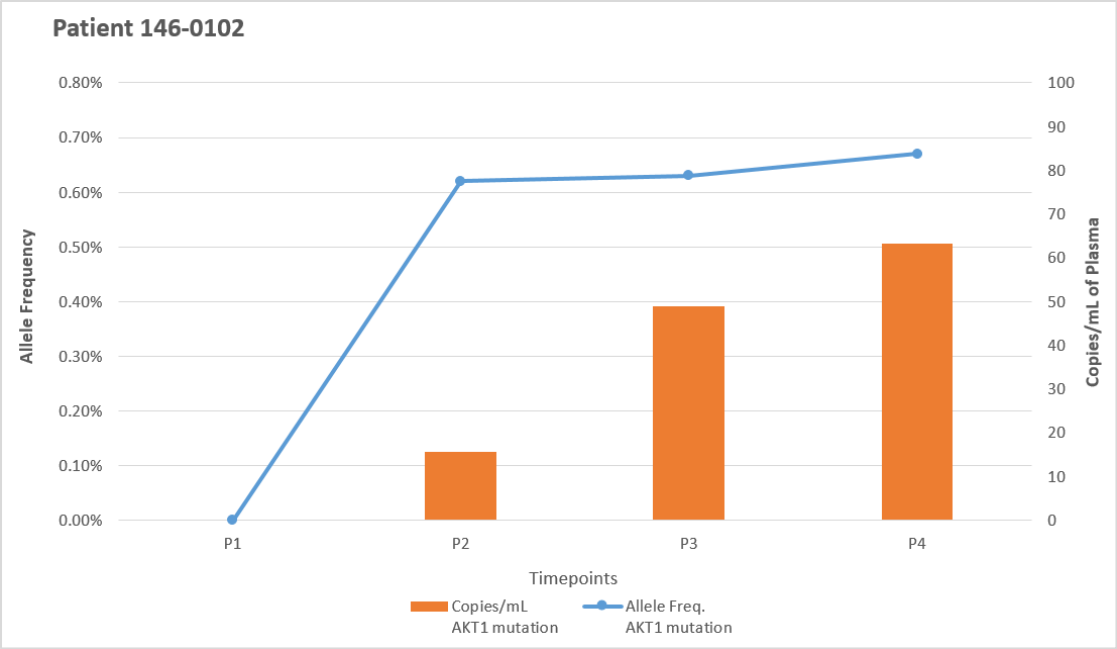


Figure 32: Longitudinal allele frequency tracking of ctDNA/ctRNA mutations in patient 146-0102. The same AKT1 mutation was observed at three of four timepoints.

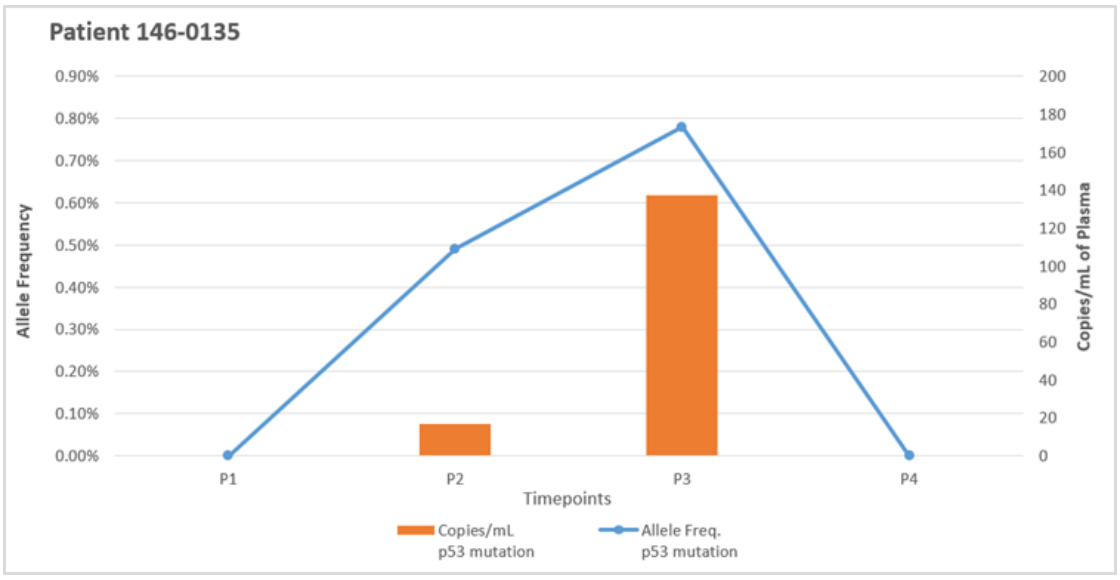


Figure 33: Longitudinal allele frequency tracking of ctDNA/ctRNA mutations in patient 146-0135. The same TP53 mutation was observed at two of four timepoints.

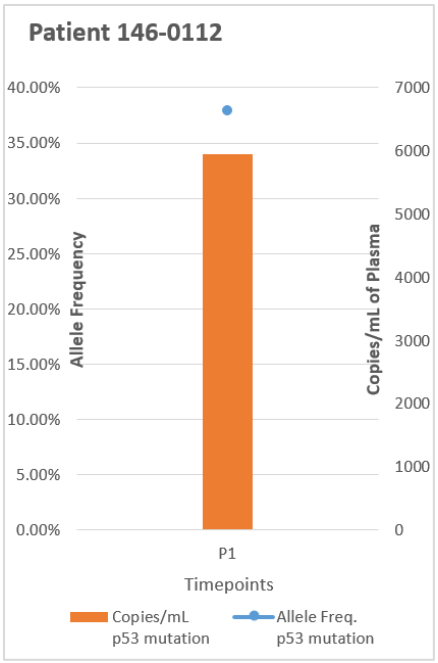


Figure 34: Longitudinal allele frequency tracking of ctDNA/ctRNA mutations in patient 146-0112. TP53 mutation was observed at the only available timepoint.

I then manually analyzed the de-duplicated BAM files on IGV to search for the same mutations identified from primary tumors in the matched plasma samples. The de-duplicated BAM files are BAM files removing those sequencing reads that might contain PCR or sequencing errors according to the molecular barcodes (Figure 35). Molecular barcodes could also be used to perform quantitative analysis to obtain copy of mutation per ml of plasma.

In this manual analysis, I was able to detect somatic mutations in the plasma of additional three patients (three TP53 mutations). Figure 36 to Figure 38 demonstrated the result of manual analysis using IGV. In patient 146-0010 (Figure 36), a TP53 mutation (Chr17:7577120, C to T) was detected in two of four timepoints. In patient 146-0014 (Figure 37), a TP53 mutation (Chr17:7579414, C to T) was detected only at timepoint 1. In patient 146-0055 (Figure 38), a TP53 mutation (Chr17:7577538, C to T) was detected in two of three timepoints. The average of DFS in those three patients is 10.8 months (4.6, 9.9, and 17.9 months; lead time: 3.23, 9.79, and 17.71 months).

In total, I was able to detect 8 of 13 patients who relapsed (62% sensitivity). The average of DFS in those 8 patients is 7.2 months. The lead time of detection of the mutation in the plasma to clinical recurrence ranged from 0.07 to 17.71 months (average: 6.71 months.)

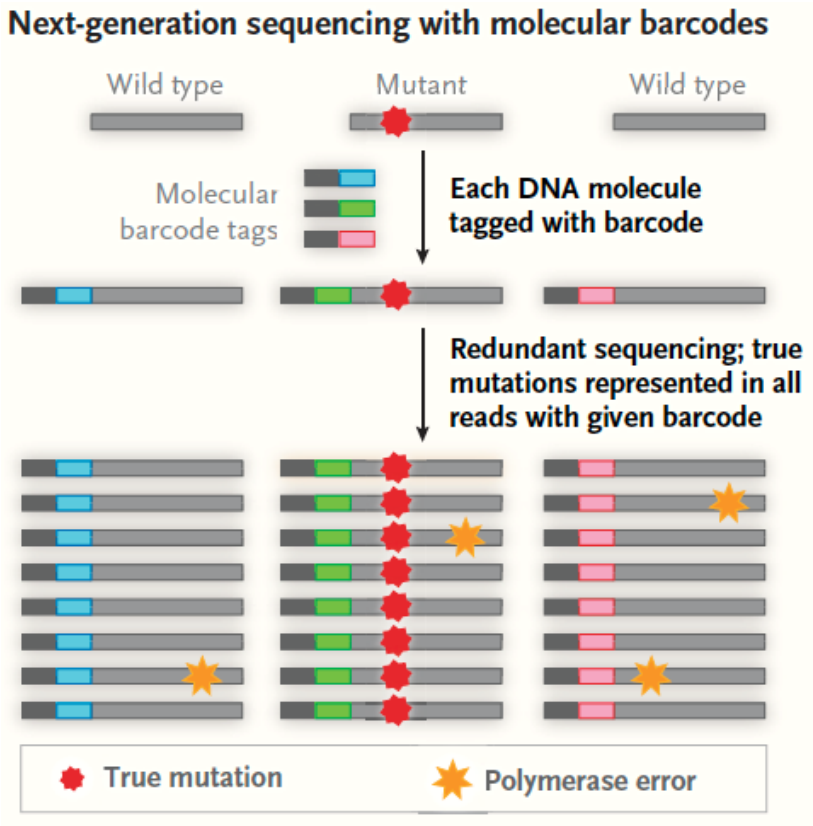


Figure 35: Concept of molecular barcode. Incorporation of molecular barcodes into each DNA molecule before PCR allows to eliminate PCR or sequencing errors during bioinfomatic analysis. Adapted from Corcoran et al. [92].

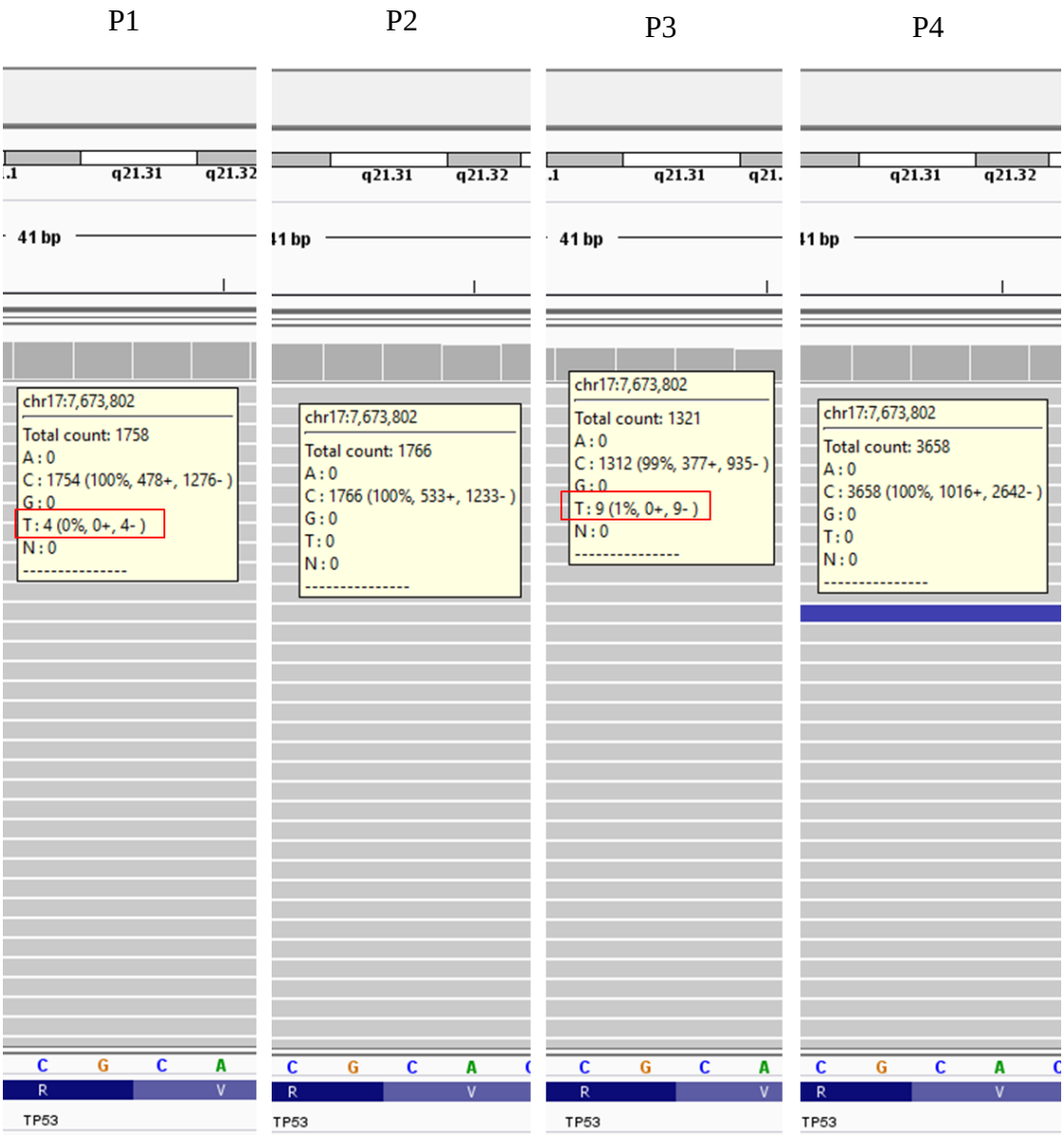


Figure 36: Manual analysis of sequencing reads containing mutation using IGV (146-0010). In patient 146-0010, TP53 mutation was detected at timepoint 1 (P1) and 3 (P3). This mutation is at locus chr17:7673802, where the wild type allele is C, whereas the mutant allele is T. Total count represents the unique sequencing reads.

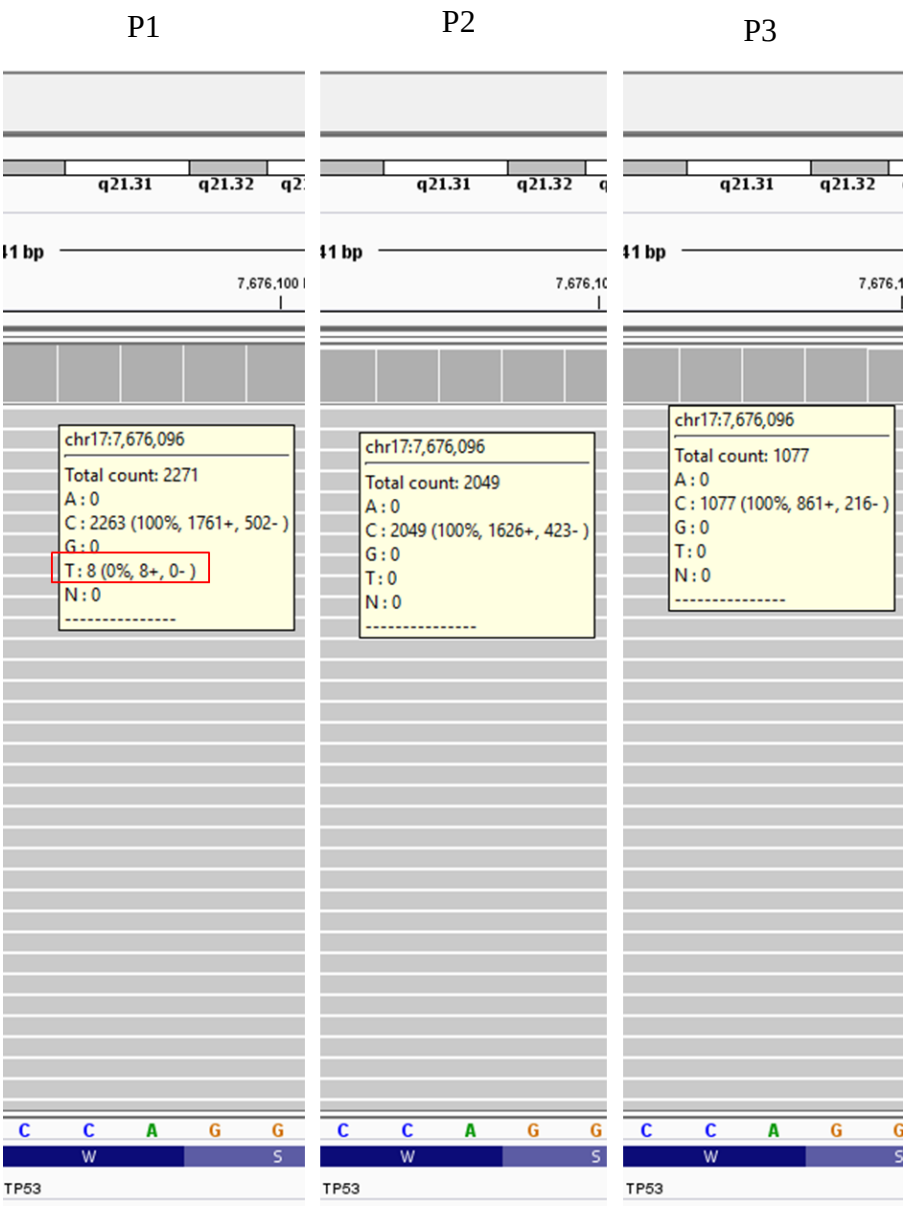


Figure 37: Manual analysis of sequencing reads containing mutation using IGV (146-0014). In patient 146-0014, TP53 mutation was detected at timepoint 1 (P1). This mutation is at locus chr17:7676096, where the wild type allele is C, whereas the mutant allele is T. Total count represents the unique sequencing reads.

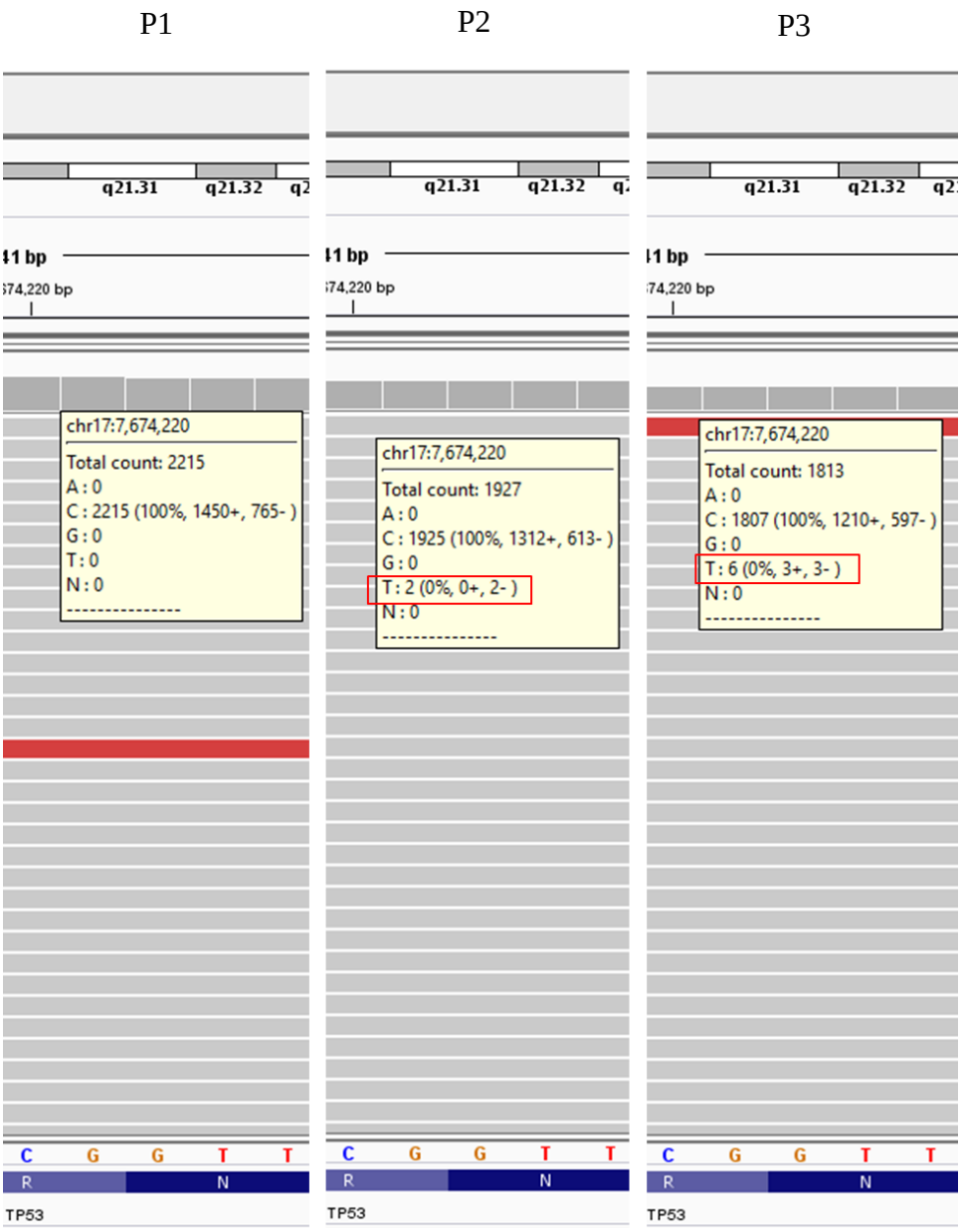


Figure 38: Manual analysis of sequencing reads containing mutation using IGV (146-0055). In patient 146-0055, TP53 mutation was detected at timepoint 2 (P2) and 3 (P3). This mutation is at locus chr17:7674220, where the wild type allele is C, whereas the mutant allele is T. Total count represents the unique sequencing reads.

3.3.3 Comparison of ctDNA-only detection versus ctDNA/ctRNA co-detection for relapse prediction

In chapter 2, I was able to detect somatic mutations in the plasma of four patients using ctDNA out of 13 who relapsed in the BRE09-146 trial. In this chapter, with incorporation of ctRNA into ctDNA sequencing assay, along with better chemistries of library preparation, I was able to increase the number of patients from four to eight with detectable mutations in the plasma. Those additional patients are: 146-0010, 146-0014, 146-0055, and 146-0135.

Another four patients (146-0005, 146-0013, 146-0102, and 146-0112) had detectable mutations using both two strategies. However, there are still some differences. In patient 146-0005, TP53 mutation was detected at two of three timepoints (P2 and P4) previously (Figure 16). Here I detected mutation at all four timepoints (Figure 30). In patient 146-0013, mutations were detected at two of four timepoints (P1 and P4 for both TP53 and CDKN2A mutations) (Figure 17). Here I detected mutations at all three timepoints for TP53; two of three timepoints for CDKN2A (P1 and P3) (Figure 31). In patient 146-0102, AKT1 mutation was detected at the only one timepoint (P1) (Figure 18). Here I detected mutation at three of four timepoints except P1 (Figure 32). In patient 146-0112, TP53 mutation was detected at the only one timepoint (P1) (Figure 19). Here I also detected mutation at the only one timepoint (P1) (Figure 34). See Table 5 for more details on allele frequency analysis using both methods.

Patient	Mutation	Timepoint	Allele Frequency	
			ctDNA	ctDNA + ctRNA
Detected in both methods				
146-0005	TP53	P1	N/A	26.76%
		P2	2.20%	11.07%
		P3	0	24.60%
		P4	49.50%	56.87%
146-0013	TP53	P1	2.70%	3.91%
		P2	0	0.12%
		P3	0	2.45%
		P4	4.80%	N/A
	CDKN2A	P1	2.00%	2.16%
		P2	0	0
		P3	0	0.77%
		P4	2.00%	N/A
146-0102	ATK1	P1	0.36%	0
		P2	N/A	0.62%
		P3	N/A	0.63%
		P4	N/A	0.67%
146-0112	TP53	P1	36.00%	37.88%
Detected only in new method				
146-0135	TP53	P1	0	0
		P2	0	0.49%
		P3	0	0.78%
		P4	0	0
146-0010	TP53	P1	0	0.23%
		P2	0	0
		P3	0	0.68%
		P4	0	0
146-0014	TP53	P1	0	0.35%
		P2	0	0
		P3	0	0
146-0055	TP53	P1	0	0
		P2	0	0.10%
		P3	0	0.33%

Table 5: Summary of sequencing result for early TNBC using ctDNA-only and ctDNA/ctRNA combination methods.

3.4 Discussion

To trace the presence of cancer using circulating materials from patients or individuals with higher risk to relapse has become very popular in recent years. Although several studies showed that some tumor-specific proteins can be potential biomarkers to detect cancers, the technologies to study novel proteins that are related to cancer are limited. In addition, the sensitivity and specificity of detecting cancers using protein biomarkers are also lacking [93, 94]. With the advances of sequencing technologies on DNA and RNA during the past decade, the paradigm of studying biomarkers from circulation has shifted to detecting circulating tumor nucleic acids using NGS and ddPCR technologies, specifically ctDNA.

In 2013, Dawson et al. showed that in metastatic breast cancer, ctDNA can confer the highest sensitivity among three most common circulating biomarkers (ctDNA, CTC, and CA 15-3) used in breast cancer detection [52]. In early detection, several studies [51, 73, 82] showed that using ctDNA had limited sensitivity in early breast cancer. A study conducted by my group in 2017 also showed congruence of moderate sensitivity in early TNBC patients who were “disease-free” during the times that the plasma samples were collected [87]. I was able to detect somatic mutations from plasma samples in 4 out of 13 relapsed patients. Details described in Chapter 2.

In this chapter, I demonstrated the strategies to resolve biological and technological issues to enhance sensitivity of detection in early TNBC. To tackle the biological limitation of using ctDNA, I first utilized ddPCR technology to validate if adding ctRNA into ctDNA assay will have more mutant molecules detected. The result

showed that in the plasma of a metastatic cancer patient, there was a 47.9% and 83.9% increase in ctDNA/ctRNA group of detecting TP53 mutation and KRAS mutation, respectively. This result showed potential application of ctDNA/ctRNA co-detection to increase the sensitivity of predicting recurrence in early breast cancer.

To enhance the technical-based sensitivity, I utilized Roche's Avenio ctDNA Analysis system, which is hybrid capture based methodology in addition to applying molecular barcode technology. As described previously, including molecular barcodes in library preparation will reduce the false positive rate by eliminating PCR and sequencing errors. Moreover, quantitative analysis can be complemented by having molecular barcode incorporated into the ends of targeted DNA molecules during the adaptor ligation step prior to any PCR reactions.

I first examined the feasibility of incorporating ctRNA into the Avenio ctDNA analysis system. In the plasma samples of six metastatic breast cancer patients, I identified the numbers of mutant molecules in the ctDNA/ctRNA co-detection group were all higher than the corresponding mutations in ctDNA-only group. The average percentage increase was 67% (range=2.8%-385%). The allele frequencies of those mutations were similar in both groups, as well as in the reports of Foundation Medicine, suggesting that similar to cfDNA, a significant amount of cfRNA was from normal tissue as well.

I then performed the same experiment on plasma samples from BRE09-146, but only for the ctDNA/ctRNA co-detection portion. I was able to detect the same mutations in those four patients with detectable ctDNA in Chapter 2 (see Table 5). In patient 146-0005, and patient 146-0013, detecting ctDNA/ctRNA showed better sensitivity: in each

mutation, at least one timepoint showed negative previously but positive in new method. On the contrary, in patient 146-0102, P1 showed positive previously but negative in new method. In patient 146-0112, it showed similar level of allele frequency in both methods.

Besides the four concordant patients, the reports showed one additional patient (146-0135) called by the standard analysis pipeline of Avenio ctDNA Analysis Server. This patient had two of four timepoints with the same TP53 mutation. Further, the manual analysis searching for the known somatic mutations revealed three additional patients with detectable mutations. Those four patients had 4.6, 6.4, 9.9, and 17.9 months of DFS individually.

To summarize this chapter, I found that by leveraging ctDNA/ctRNA co-detection (improving the biological-based sensitivity), along with better NGS technology (improving the technical-based sensitivity), I was able to detect somatic mutations in plasma samples from 8 out of 13 patients who had a recurrence in early TNBC. The sensitivity of MRD detection was improved from 31% to 62%. The average DFS was increased from 4.6 months to 7.2 months. The most distant recurrence detected was 17.9 months. The lead time of detection of the mutation in the plasma to clinical recurrence ranged from 0.07 to 17.71 months (average: 6.71 months). Since I only processed plasma samples from patients who had a recurrence so far, only sensitivity is available. The rest of the plasma samples from patients who had no recurrence will need to be processed in the near future to obtain specificity.

Chapter 4: Summary

In the recent years, studies indicated that some of the TNBC patients may benefit from one of two FDA-approved treatments other than chemotherapy: (1) PARP inhibitor Talazoparib for patients with advanced or metastatic HER2-negative breast cancer and germline BRCA1/2 mutations [12, 13]; (2) immunotherapeutic agent Atezolizumab for unresectable locally advanced or metastatic TNBC patients with PD-L1 positive tumors [14]. Although there have been advances in therapies for TNBC, the current standard of care for the majority of TNBC patients, in early and advanced-stages, is still chemotherapy. A significant proportion of patients with early-stage TNBC are treated with neoadjuvant chemotherapy. A stark dichotomy exists in outcome based on response to neoadjuvant therapy. Approximately a third of patients will achieve a pCR and will have a favorable overall survival outcome. In contradistinction, two-thirds of patients will have RD after neoadjuvant chemotherapy and are at a high risk of relapse. For the TNBC patients with RD, detecting the presence of MRD earlier is critical to be able to intervene and perhaps increase the survival.

Precision medicine has become widely accepted lately, especially in oncology, for the diagnosis, treatment selection, and therapy observation by utilizing patient's molecular profiling. Tissue biopsies are the standard for cancer diagnosis for molecular profiling to guide the selection of therapy. However, there are drawbacks of using tissue biopsies. For example, one single biopsy cannot completely represent the heterogeneity of the whole tumor or metastases. In addition, some lesions are not safe or feasible to access. Another important tool of precision medicine is liquid biopsies, an emerging

method for minimally-invasive cancer detection by analyzing the tumor-derived material from bodily fluids, mainly blood. Liquid biopsies have the advantages in clinical application for several reasons: (1) they are considered non- or minimal-invasive; (2) serial sampling allows clinicians to observe the real-time status of patient instead of one snapshot; (3) compared to standard diagnosis tools, such as imaging which has strict regulations for the frequency. Liquid biopsies can be performed repeatedly over the course of therapy.

Blood-based cancer biomarkers can be defined as any biological materials released by tumor tissues. Previously, protein biomarkers were widely used in cancer diagnostics within liquid biopsies. There are two protein biomarkers approved by FDA which are more relevant to breast cancer: carcinoembryonic antigen (CEA) and MUC-1 (including MCA, BRMA CA549, CA27.29, and CA 15-3) [94]. However, the major disadvantages of these two protein biomarkers are lack of sensitivity and specificity [93]. In 2013, Dawson et al. compared three blood-based biomarkers in metastatic patients: CA 15-3, CTCs, and ctDNA. The result indicated that ctDNA has the highest sensitivity among those three biomarkers [52]. The origin of ctDNA is from tumor tissues or CTCs. Because ctDNA carries tumor-specific aberrations, it is considered to be a more favorable biomarker over a protein biomarker. Although ctDNA usually comprises very small portion of total cfDNA in early-stage cancer patients, with the advances of molecular technologies I are able to detect the low amount of ctDNA out of total cfDNA.

In this dissertation, I first showed that by detecting ctDNA, I was able to predict rapid recurrence, but not late recurrence in early TNBC after neoadjuvant chemotherapy. The sensitivity of detection was moderate and with high specificity.

To tackle the biological-based issue I incorporated ctRNA into ctDNA detection. Unlike ctDNA, ctRNA can be actively secreted from living tumor cells within exosomes. I expected to see an increase in mutant molecules detected after adding ctRNA into ctDNA. My preliminary data of ddPCR applied on a single metastatic cancer patient supported my hypothesis. The ddPCR technology is considered to have a higher sensitivity compared to NGS technology. However, ddPCR can usually only screen one variant at a time. Moreover, in my study, several mutations were located at the edge of exonic regions, making it impossible to design an assay for both ctDNA and ctRNA detection simultaneously. Therefore, NGS technology is still the best fit in this study for multiple-mutation detection and ctDNA/ctRNA co-detection.

I also worked on finding a better solution to enhance the technical-based sensitivity. In chapter 2, I utilized amplicon-based technology to detect ctDNA. In the analysis, I found that there were significant amounts of false positive calls by comparing the results to corresponding tumor sequencing data. In chapter 3, I utilized the Roche Avenio ctDNA analysis system, which applies hybrid capture enrichment along with molecular barcode technology to eliminate false positive signals. It is also considered to detect ctDNA down to 0.1% allele frequency. Although the Roche Avenio ctDNA analysis kit was designed for ctDNA detection only, I successfully merged ctRNA detection into the existing system by conducting the proof of concept experiment on plasma samples from six metastatic breast cancer patients. The results showed that comparing to ctDNA-only, ctDNA/ctRNA co-detection can increase the detectable mutant molecules from 2.8% to 385%. By improving biological-based and technical-based sensitivity, I was able to enhance the sensitivity of detection from 31% to 62% in

early TNBC patients who were considered disease-free during plasma collections. I increased the range maximum of lead time from 8.87 to 17.71 months. The specificity of new strategy will be confirmed by the following experiments in the future. For patients who had a recurrence but did not have detectable mutant molecules from their plasma samples, most had a late recurrence (DFS range: 22.17 to 33.74 months) except for two patients who had a rapid recurrence (2.66 and 3.12 months DFS). The possible explanation for these two patients without detectable mutation could be due to the rapid progression in a very short time of their plasma collection. This dissertation also presents a significant look at detecting ctDNA-only (amplicon-based) versus ctDNA/ctRNA (hybrid capture) as the samples collected from patients were split into two aliquots to allow for comparison.

Other studies [51, 82] showed serial sampling can increase the detection of sensitivity over time. However, my samples were from a retrospective study, which had plasma samples collected in the four months after neoadjuvant chemotherapy and surgery. This made serial sampling impossible under current setting. Although there are limitations based on this timeline of samples, my purpose was to detect the MRD in the very early stage of TNBC to predict who will relapse. Another limitation of my study was the amount of plasma as most of my samples only had one milliliter available.

In summation, this dissertation provides the first observation for co-detection of ctDNA and ctRNA in the setting of early TNBC patients who were considered “disease-free” after neoadjuvant chemotherapy. I successfully enhanced the sensitivity of detection from 31% to 62%. With this work, I hope that it can provide early TNBC patients to have better opportunities for more timely therapeutic intervention and eventually improve the

overall survival. The future directions for further improving the detection of MRD can be focused on two parts: (1) increasing the amount of plasma collection; (2) incorporating other biomarkers for co-detection, such as microRNA, lncRNA, methylation, and/or protein.

REFERENCES

1. National Institutes of Health, National Cancer Institute. *Surveillance, Epidemiology, and End Results Program. Cancer stat facts: female breast cancer.* 2019 [cited 2019 April 2]; Available from: https://seer.cancer.gov/statfacts/html/breast.html.
2. Anders, C. and L.A. Carey, *Understanding and treating triple-negative breast cancer.* Oncology (Williston Park), 2008. **22**(11): p. 1233-9; discussion 1239-40, 1243.
3. Irvin, W.J., Jr. and L.A. Carey, *What is triple-negative breast cancer?* Eur J Cancer, 2008. **44**(18): p. 2799-805.
4. Pal, S.K. and J. Mortimer, *Triple-negative breast cancer: novel therapies and new directions.* Maturitas, 2009. **63**(4): p. 269-74.
5. Rivenbark, A.G., S.M. O'Connor, and W.B. Coleman, *Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine.* Am J Pathol, 2013. **183**(4): p. 1113-1124.
6. Tomao, F., et al., *Triple-negative breast cancer: new perspectives for targeted therapies.* Onco Targets Ther, 2015. **8**: p. 177-93.
7. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clin Cancer Res, 2007. **13**(15 Pt 1): p. 4429-34.
8. Carotenuto, P., et al., *Triple negative breast cancer: from molecular portrait to therapeutic intervention.* Crit Rev Eukaryot Gene Expr, 2010. **20**(1): p. 17-34.
9. Lehmann, B.D. and J.A. Pietenpol, *Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes.* J Pathol, 2014. **232**(2): p. 142-50.
10. Foulkes, W.D., I.E. Smith, and J.S. Reis-Filho, *Triple-negative breast cancer.* N Engl J Med, 2010. **363**(20): p. 1938-48.
11. Wetterskog, D., et al., *Adenoid cystic carcinomas constitute a genomically distinct subgroup of triple-negative and basal-like breast cancers.* J Pathol, 2012. **226**(1): p. 84-96.
12. *Talazoparib Bests Chemo for Breast Cancer.* Cancer Discov, 2018. **8**(2): p. OF3.
13. Ettl, J., et al., *Quality of life with talazoparib versus physician's choice of chemotherapy in patients with advanced breast cancer and germline BRCA1/2 mutation: patient-reported outcomes from the EMBRACA phase III trial.* Ann Oncol, 2018. **29**(9): p. 1939-1947.
14. Schmid, P., et al., *Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer.* N Engl J Med, 2018. **379**(22): p. 2108-2121.
15. Liedtke, C., et al., *Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer.* J Clin Oncol, 2008. **26**(8): p. 1275-81.
16. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
17. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
18. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* JAMA, 2006. **295**(21): p. 2492-502.

19. Foulkes, W.D., et al., *Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer.* J Natl Cancer Inst, 2003. **95**(19): p. 1482-5.
20. Kruger, K., et al., *Expression of Nestin associates with BRCA1 mutations, a basal-like phenotype and aggressive breast cancer.* Sci Rep, 2017. **7**(1): p. 1089.
21. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
22. Perou, C.M., *Molecular stratification of triple-negative breast cancers.* Oncologist, 2010. **15 Suppl 5**: p. 39-48.
23. Prat, A., et al., *Predicting response and survival in chemotherapy-treated triple-negative breast cancer.* Br J Cancer, 2014. **111**(8): p. 1532-41.
24. Prat, A., et al., *Molecular characterization of basal-like and non-basal-like triple-negative breast cancer.* Oncologist, 2013. **18**(2): p. 123-33.
25. van der Groep, P., et al., *Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data.* J Clin Pathol, 2006. **59**(6): p. 611-7.
26. Antoniou, A., et al., *Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies.* Am J Hum Genet, 2003. **72**(5): p. 1117-30.
27. Reis-Filho, J.S. and A.N. Tutt, *Triple negative tumours: a critical review.* Histopathology, 2008. **52**(1): p. 108-18.
28. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Basal-like breast cancer: a critical review.* J Clin Oncol, 2008. **26**(15): p. 2568-81.
29. Wan, J.C.M., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA.* Nat Rev Cancer, 2017. **17**(4): p. 223-238.
30. Siravegna, G., et al., *Integrating liquid biopsies into the management of cancer.* Nat Rev Clin Oncol, 2017. **14**(9): p. 531-548.
31. Chan, K.C., et al., *Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing.* Proc Natl Acad Sci U S A, 2013. **110**(47): p. 18761-8.
32. Amorim, M.G., et al., *A total transcriptome profiling method for plasma-derived extracellular vesicles: applications for liquid biopsies.* Sci Rep, 2017. **7**(1): p. 14395.
33. Cohen, J.D., et al., *Detection and localization of surgically resectable cancers with a multi-analyte blood test.* Science, 2018. **359**(6378): p. 926-930.
34. Cohen, J.D., et al., *Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers.* Proc Natl Acad Sci U S A, 2017. **114**(38): p. 10202-10207.
35. Kim, Y., et al., *Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer.* Nat Commun, 2016. **7**: p. 11906.
36. Mayers, J.R., et al., *Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development.* Nat Med, 2014. **20**(10): p. 1193-1198.
37. Sasaroli, D., G. Coukos, and N. Scholler, *Beyond CA125: the coming of age of ovarian cancer biomarkers. Are we there yet?* Biomark Med, 2009. **3**(3): p. 275-288.

38. Nossov, V., et al., *The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125?* Am J Obstet Gynecol, 2008. **199**(3): p. 215-23.
39. Duffy, M.J., D. Evoy, and E.W. McDermott, *CA 15-3: uses and limitation as a biomarker for breast cancer.* Clin Chim Acta, 2010. **411**(23-24): p. 1869-74.
40. Adhyam, M. and A.K. Gupta, *A Review on the Clinical Utility of PSA in Cancer Prostate.* Indian J Surg Oncol, 2012. **3**(2): p. 120-9.
41. Mandel, P. and P. Metais, *[Not Available]*. C R Seances Soc Biol Fil, 1948. **142**(3-4): p. 241-3.
42. Leon, S.A., et al., *Free DNA in the serum of cancer patients and the effect of therapy.* Cancer Res, 1977. **37**(3): p. 646-50.
43. Nawroz, H., et al., *Microsatellite alterations in serum DNA of head and neck cancer patients.* Nat Med, 1996. **2**(9): p. 1035-7.
44. Chen, X.Q., et al., *Microsatellite alterations in plasma DNA of small cell lung cancer patients.* Nat Med, 1996. **2**(9): p. 1033-5.
45. Sorenson, G.D., et al., *Soluble normal and mutated DNA sequences from single-copy genes in human blood.* Cancer Epidemiol Biomarkers Prev, 1994. **3**(1): p. 67-71.
46. Vasioukhin, V., et al., *Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia.* Br J Haematol, 1994. **86**(4): p. 774-9.
47. Lo, Y.M., et al., *Presence of fetal DNA in maternal plasma and serum.* Lancet, 1997. **350**(9076): p. 485-7.
48. Wong, H.Y. and B.H. Park, *Plasma tumor DNA: on your markers, get set, go!* Ann Transl Med, 2014. **2**(1): p. 2.
49. Shaw, J.A. and J. Stebbing, *Circulating free DNA in the management of breast cancer.* Ann Transl Med, 2014. **2**(1): p. 3.
50. Heidary, M., et al., *The dynamic range of circulating tumor DNA in metastatic breast cancer.* Breast Cancer Res, 2014. **16**(4): p. 421.
51. Garcia-Murillas, I., et al., *Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer.* Sci Transl Med, 2015. **7**(302): p. 302ra133.
52. Dawson, S.J., et al., *Analysis of circulating tumor DNA to monitor metastatic breast cancer.* N Engl J Med, 2013. **368**(13): p. 1199-209.
53. Schutz, E., et al., *Chromosomal instability in cell-free DNA is a serum biomarker for prostate cancer.* Clin Chem, 2015. **61**(1): p. 239-48.
54. Hamakawa, T., et al., *Monitoring gastric cancer progression with circulating tumour DNA.* Br J Cancer, 2015. **112**(2): p. 352-6.
55. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood.* Nat Rev Clin Oncol, 2013. **10**(8): p. 472-84.
56. Xu, J.F., et al., *Exosomes containing differential expression of microRNA and mRNA in osteosarcoma that can predict response to chemotherapy.* Oncotarget, 2017. **8**(44): p. 75968-75978.
57. Prendergast, E.N., et al., *Optimizing exosomal RNA isolation for RNA-Seq analyses of archival sera specimens.* PLoS One, 2018. **13**(5): p. e0196913.

58. Hydbring, P., et al., *Exosomal RNA-profiling of pleural effusions identifies adenocarcinoma patients through elevated miR-200 and LCN2 expression.* Lung Cancer, 2018. **124**: p. 45-52.
59. Dong, L., et al., *Circulating Long RNAs in Serum Extracellular Vesicles: Their Characterization and Potential Application as Biomarkers for Diagnosis of Colorectal Cancer.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(7): p. 1158-66.
60. Mattioli, K., et al., *High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity.* Genome Res, 2019. **29**(3): p. 344-355.
61. Lin, L.Y., et al., *Tumor-originated exosomal lncUEGC1 as a circulating biomarker for early-stage gastric cancer.* Mol Cancer, 2018. **17**(1): p. 84.
62. Kern, C., et al., *Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species.* BMC Genomics, 2018. **19**(1): p. 684.
63. Do Canto, L.M., et al., *MicroRNA analysis of breast ductal fluid in breast cancer patients.* Int J Oncol, 2016. **48**(5): p. 2071-8.
64. Clark, E.A., et al., *Concise review: MicroRNA function in multipotent mesenchymal stromal cells.* Stem Cells, 2014. **32**(5): p. 1074-82.
65. Hartmaier, R.J., et al., *Recurrent hyperactive ESR1 fusion proteins in endocrine therapy-resistant breast cancer.* Ann Oncol, 2018. **29**(4): p. 872-880.
66. Malcher, C., et al., *Development of a comprehensive noninvasive prenatal test.* Genet Mol Biol, 2018. **41**(3): p. 545-554.
67. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage.* Nat Med, 2014. **20**(5): p. 548-54.
68. Niu, C., et al., *Ultrasensitive Single Fluorescence-Labeled Probe-Mediated Single Universal Primer-Multiplex-Droplet Digital Polymerase Chain Reaction for High-Throughput Genetically Modified Organism Screening.* Anal Chem, 2018. **90**(9): p. 5586-5593.
69. Ono, Y., et al., *An improved digital polymerase chain reaction protocol to capture low-copy KRAS mutations in plasma cell-free DNA by resolving 'subsampling' issues.* Mol Oncol, 2017. **11**(10): p. 1448-1458.
70. Tan, C., et al., *A multiplex droplet digital PCR assay for non-invasive prenatal testing of fetal aneuploidies.* Analyst, 2019. **144**(7): p. 2239-2247.
71. Vidal-Folch, N., et al., *A Droplet Digital PCR Method for Severe Combined Immunodeficiency Newborn Screening.* J Mol Diagn, 2017. **19**(5): p. 755-765.
72. Yu, Q., et al., *Multiplex picoliter-droplet digital PCR for quantitative assessment of EGFR mutations in circulating cell-free DNA derived from advanced non-small cell lung cancer patients.* Mol Med Rep, 2017. **16**(2): p. 1157-1166.
73. Riva, F., et al., *Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer.* Clin Chem, 2017. **63**(3): p. 691-699.
74. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
75. Diehl, F., et al., *Detection and quantification of mutations in the plasma of patients with colorectal tumors.* Proc Natl Acad Sci U S A, 2005. **102**(45): p. 16368-73.

76. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
77. Holdhoff, M., et al., *Analysis of circulating tumor DNA to confirm somatic KRAS mutations.* J Natl Cancer Inst, 2009. **101**(18): p. 1284-5.
78. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.
79. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
80. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
81. Masuda, N., et al., *Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy.* N Engl J Med, 2017. **376**(22): p. 2147-2159.
82. Olsson, E., et al., *Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease.* EMBO Mol Med, 2015. **7**(8): p. 1034-47.
83. Gootenberg, J.S., et al., *Nucleic acid detection with CRISPR-Cas13a/C2c2.* Science, 2017. **356**(6336): p. 438-442.
84. Melo, S.A., et al., *Glypican-1 identifies cancer exosomes and detects early pancreatic cancer.* Nature, 2015. **523**(7559): p. 177-82.
85. Ma, F., et al., *ctDNA dynamics: a novel indicator to track resistance in metastatic breast cancer treated with anti-HER2 therapy.* Oncotarget, 2016. **7**(40): p. 66020-66031.
86. Murtaza, M., et al., *Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer.* Nat Commun, 2015. **6**: p. 8760.
87. Chen, Y.H., et al., *Next-generation sequencing of circulating tumor DNA to predict recurrence in triple-negative breast cancer patients with residual disease after neoadjuvant chemotherapy.* NPJ Breast Cancer, 2017. **3**: p. 24.
88. Chaudhuri, A.A., et al., *Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling.* Cancer Discov, 2017. **7**(12): p. 1394-1403.
89. Newman, A.M., et al., *Integrated digital error suppression for improved detection of circulating tumor DNA.* Nat Biotechnol, 2016. **34**(5): p. 547-555.
90. Enderle, D., *Highly sensitive detection of low abundant somatic mutations in circulating exosomal RNA and cfDNA with next-generation sequencing.* American Society of Clinical Oncolgy 2015 Annual Meeting Abstract #11061, 2015.
91. Enderle, D.K., T.; Sullivan, R. J.; Speil, A.; Brinkmann, K.; Mueller, R.; Bentink, S.; O'Neill, V.; Skog, J.; Noerholm, M.; Berking, C.; Flaherty, K. T., *Monitoring therapy response and resistance mutations in circulating RNA and DNA of plasma from melanoma patients.* American Society of Clinical Oncolgy 2015 Annual Meeting Poster Presentation, 2015.
92. Corcoran, R.B. and B.A. Chabner, *Application of Cell-free DNA Analysis to Cancer Treatment.* N Engl J Med, 2018. **379**(18): p. 1754-1765.

93. Duffy, M.J., *Serum tumor markers in breast cancer: are they of clinical value?* Clin Chem, 2006. **52**(3): p. 345-51.
94. Gam, L.H., *Breast cancer and protein biomarkers.* World J Exp Med, 2012. **2**(5): p. 86-91.

CURRICULUM VITAE

Yu-Hsiang Chen

Education

Graduate: (1) Indiana University, IUPUI, Indianapolis, IN (June 2019)
Major: Medical & Molecular Genetics, Doctor of Philosophy
Minor: Cancer Biology
(2) Purdue University, IUPUI, Indianapolis, IN (July 2013)
Major: Biology, Master of Science

Undergraduate: National Taiwan Normal University, Taipei, Taiwan (June 2006)
Major: Life Science, Bachelor of Science

Honors and Awards

1) AACR San Antonio Breast Cancer Symposium 2015 Susan G. Komen®Scholar-in-Training Award.
2) Purdue University 2016 Big Data Training for Translational Omics Research NIH-funded Boot Camp.
3) Indiana University School of Medicine 2016 Annual Cancer Research Day – Honorable Mention.
4) Indiana University School of Medicine 2017 Graduate Student Travel Grant.

Previous Scientific Work Experience

1) Topic Research Undergraduate. National Taiwan Normal University, Taiwan. July 2004 – June 2006.

2) Bioinformatics Program Research Student. National Taiwan University, Taiwan. September 2005 – June 2006.
3) Research Assistant / Lead of Illumina Team. High-Throughput Genomics Core Facility, Biodiversity Research Center, Academia Sinica, Taiwan. August 2008 – June 2011.

Meeting Presentations

1) **YH Chen**, A Malkova. (2012, May). *Genomic instability during break-induced replication in yeast*. Poster presented at 14th Annual Midwest DNA Repair Symposium, Cincinnati, OH.
2) **YH Chen**, BA Hancock, JP Solzak, KD Miller, M Radovich. (2015, December). *Next generation sequencing of circulating tumor DNA to predict recurrence in triple-negative breast cancer patients with residual disease after neoadjuvant chemotherapy.* Poster presented at the annual meeting of the San Antonio Breast Cancer Symposium, San Antonio, TX.
3) BA Hancock, **YH Chen**, JP Solzak, KD Miller, M Radovich. (2016, April). *TP53 mutation is a biomarker for prognosis in triple-negative breast cancer patients treated with post-neoadjuvant cisplatin.* Poster presented at the annual meeting of the American Association for Cancer Research, New Orleans, LA.
4) BA Hancock, **YH Chen**, JP Solzak, MN Ahmad, DC Wedge, D Brinza, C Scafe, JVeitch, R Gottimukkala, W Short, RV Atale, M Ivan, SS Badve, BP Schneider, KD Miller, M Radovich. (2017, December). *Molecular regulators of resistance and relapse in chemorefractory triple-negative breast cancers.*

Poster presented at the annual meeting of the San Antonio Breast Cancer Symposium, San Antonio, TX.

5) **YH Chen**, BA Hancock, JP Solzak, M Radovich. (2019, April). *Co-detection of circulating tumor DNA and RNA in the plasma of patients with breast cancer increases the detectable number of mutated molecules.* Poster presented at the annual meeting of the American Association of Cancer Research, Atlanta, GA.

Publications

1) YC Chang, CY Lin, CM Hsu, HC Lin, **YH Chen**, GJ Lee-Chen, MT Su, LS Ro, CM Chen HM Hsieh-Li (2011) "Neuroprotective Effects of Granulocyte-colony Stimulating Factor in a Navel Transgenic Mouse Model of SCA17", Journal of Neurochemistry 2011, 118(2):288-303.
2) TY Wang, HL Chen, MY Lu, YC Chen, HM Sung, CT Mao, HY Cho, HM Ke, TY Hwa, SK Ruan, KY Hung, CK Chen, JY Li, YC Wu, **YH Chen**, SP Chou, YW Tsai, TC Chu, CC A Shih, WS Li, MC Shih (2011) "Functional Characterization of Cellulases Identified from the cow rumen fungus Neocallimastix Patriciarum W5 by Transcriptomic and Secretomis Analyses", Biotechnology for Biofuels 2011, 4:24.
3) WL Fan, CS Ng, CF Chen, MY Lu, **YH Chen**, CJ Liu, SM Wu, CK Chen, JJ Chen, CT Mao, YT Lai, WS Lo, WH Chang, WH Li (2013) "Genome-wide Patterns of Genetics Variation in Two Domestic Chickens", Genome Biology and Evolution 2013, 5(7):1376-92.

4) HY Gao, T Hawkins, A Jasti, **YH Chen**, K Mockaitis, M Dinauer, K Cornetta (2014) “Development and Evaluation of Quality Metrics for Bioinformatics Analysis of Viral Insertion Site Data Generated Using High Throughput Sequencing”, Biomedicines, 2(2):195-210.

5) E Sasaki, X Zhang, HG Sun, MY Lu, TL Liu, A Ou, JY Li, **YH Chen**, SE Ealick, HW Liu (2014) “Co-opting Sulphur-Carrier Proteins from Primary Metabolic Pathways for 2-thiosugar Biosynthesis”, Nature 2014, 510(7505):427-31.

6) **YH Chen**, BA Hancock, JP Solzak, D Brinza, C Scafe, KD Miller, M Radovich (2017) “Next-generation Sequencing of Circulating Tumor DNA to Predict Recurrence in Triple-negative Breast Cancer Patients with Residual Disease after Neoadjuvant Chemotherapy”, NPJ Breast Cancer 2017, 3:24. doi:10.1038/s41523-017-0028-4